

**$t\bar{t}$ Production Cross Section Measurement using
Soft Electron Tagging in $p\bar{p}$ Collisions at
 $\sqrt{s} = 1.96$ TeV**

A thesis presented

by

John Paul Chou

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

September 2008

©2008 - John Paul Chou

All rights reserved.

Thesis advisor

Author

Melissa Franklin

John Paul Chou

$t\bar{t}$ Production Cross Section Measurement using Soft Electron Tagging in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV

Abstract

We measure the production cross section of $t\bar{t}$ events in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The data was collected by the CDF experiment in Run 2 of the Tevatron accelerator at the Fermi National Accelerator Laboratory between 2002 and 2007. 1.7 fb^{-1} of data was recorded during this time period. We reconstruct $t\bar{t}$ events in the *lepton+jets* channel, whereby one W boson - resulting from the decay of the top quark pairs - decays leptonically and the other hadronically. The dominant background to this process is the production of W bosons in association with multiple jets. To distinguish $t\bar{t}$ from background, we identify soft electrons from the semileptonic decay of heavy flavor jets produced in $t\bar{t}$ events. We measure a cross section of $\sigma_{p\bar{p}} = 7.8 \pm 2.4(\text{stat}) \pm 1.6(\text{syst}) \pm 0.5(\text{lumi})$.

Contents

Title Page	i
Abstract	iii
Table of Contents	iv
List of Figures	viii
List of Tables	xviii
Acknowledgments	xxii
Dedication	xxiv
1 Physics of The Standard Model and Beyond	1
1.1 The Standard Model	2
1.2 The Top Quark	4
1.2.1 Expectation	4
1.2.2 Mass	6
1.2.3 Hadroproduction	8
Pair Production	8
Parton Distribution Functions	10
Renormalization/Factorization Scale	11
Single Top Production	13
1.2.4 Decay	14
1.3 Soft Electron Tagging Measurements	16
1.3.1 $t\bar{t}$ Cross Section	16
1.3.2 Production of W plus Charm	16
2 Experimental Apparatus	19
2.1 The Accelerator Complex	19
2.1.1 Proton Acceleration: From Source to Main Injector	19
2.1.2 Antiproton Production	21
2.1.3 Tevatron	22
2.2 The Collider Detector at Fermilab	23
2.2.1 Tracking	25
Silicon	26

	Drift Chamber	29
2.2.2	Calorimetry	30
2.2.3	Muon Chambers	32
2.2.4	Luminosity Monitors	34
2.2.5	Trigger and Data Acquisition Systems	35
3	Event Reconstruction	38
3.1	Tracks	38
3.2	Leptons	40
3.2.1	Electrons	40
3.2.2	Muons	43
3.3	Calorimeter Objects	45
3.3.1	Jets	45
3.3.2	Missing Energy	47
3.4	b -Tagging	48
3.4.1	Beamline and Primary Vertex Reconstruction	48
3.4.2	Secondary Vertex Reconstruction	49
4	Soft Electron Tagging	51
4.1	Introduction	51
4.2	Sample Selection	56
4.2.1	Pure Conversion Selection	56
4.2.2	Generic Track Selection: Fake Electrons	63
4.3	Electron Identification Variables	64
4.3.1	Calorimeter Variables	65
4.3.2	Shower Max Clustering	71
4.3.3	χ^2 Variables	76
4.3.4	Δ_{wire} and Δ_{strip} Variables	77
4.3.5	Likelihood Construction	81
4.4	Tagger Definition	85
4.5	Tagging Efficiency	87
5	Conversion Identification	94
5.1	Introduction	94
5.2	Sample Selection	98
5.2.1	Electron Tag Selection	99
5.2.2	Anti-Electron Tag Selection	99
5.2.3	Pure Conversion Electron Selection	101
	Sample Construction	101
	Sample Purity	101
5.3	Missing Silicon Layers Description	104
5.4	Conversion Over-Efficiency	106

5.4.1	Over-Efficiency Measurement	106
5.4.2	Over-Efficiency Systematics	107
5.5	Conversion Efficiency	109
5.5.1	Partner Matching Efficiency Measurement	110
5.5.2	Missing Silicon Layer Efficiency Measurement	112
5.5.3	Combined Efficiency Measurement	113
5.6	Method Cross Check	116
5.7	Scale Factors	117
5.7.1	Efficiency Scale Factor	117
5.7.2	Over-Efficiency Scale Factor	120
6	Tag Matrix	123
6.1	Tag Matrix Definition	124
6.2	High p_T Extrapolation	126
6.3	Conversion Sample Cross Check	128
6.4	Calorimeter Modeling	129
7	Fake Matrix	135
7.1	Fake Matrix Definition	136
7.2	Electron Contamination	154
7.2.1	Conversion Electron Contamination	154
7.2.2	Heavy Flavor Electron Contamination	156
7.2.3	Dalitz and Other Electron Contamination	158
7.2.4	Total Electron Contamination	159
7.3	Cross Checks and Systematics	162
7.4	Summary	164
8	SLT_e Cross Check in $b\bar{b}$ Events	167
8.1	$b\bar{b}$ Sample Selection	168
8.2	Application of the SLT_e Tagger	171
8.3	Data/MC Comparison	172
9	Cross Section	194
9.1	Event Selection	194
9.2	Signal Expectation	197
9.3	Background Expectation	200
9.3.1	MC Derived Backgrounds	201
9.3.2	QCD	205
9.3.3	W+Jets	208
	HF Fraction Estimation and Calibration	209
	Overlap Removal Scheme	211
9.4	Optimization	213

9.5 Measurement	214
10 Conclusions	231
Bibliography	234

List of Figures

1.1	Feynman diagram showing the SM tree order process of $B_s \rightarrow \mu^+ \mu^-$.	4
1.2	Triangle diagram, showing the anomalous term. The diagrams needed to show that the electroweak theory is free of anomalies has three external gauge bosons, and the internal lines are all possible fermions.	5
1.3	Feynman diagram showing radiative corrections from the top quark and Higgs to the W mass.	7
1.4	Higgs mass expected from precision measurements of the top and W mass. Shown are the theoretical regions favored by a purely SM Higgs and a Higgs in the MSSM. Assumed values: $M_{top} = 172.6 \pm 1.4 \text{ GeV}/c^2$ and $M_W = 80.398 \pm 0.025 \text{ GeV}/c^2$.	7
1.5	Feynman diagrams of quark annihilation (left) and gluon fusion (right). These are the two prominent production mechanisms of top at the Tevatron.	8
1.6	$t\bar{t}$ production cross section as a function of the top mass. Shown is the combined CDF Run 2 result with 760 pb^{-1} .	9
1.7	$t\bar{t}$ production cross section as a function of the \sqrt{s} energy. Shown are the CDF Run 1 and CDF Run 2 results.	10
1.8	Next-to-leading order PDFs for protons from the CTEQ6M parameterization, where x is the momentum fraction, and $xf(x)$ is the probability density. $Q = 2$ and 100 GeV on the left and right, respectively.	11
1.9	Feynman diagrams showing single top production in the s-channel (left) and t-channel (right).	13
1.10	Feynman diagram of tree level $t\bar{t}$ production and subsequent lepton+jets decay. The lepton and neutrino have been highlighted to note the decay channel.	15
1.11	Combined results from CDF of the $t\bar{t}$ production cross section. Results shown include the cross section result from this measurement.	17
1.12	Tree order diagram of W hadroproduction associated with a single charm quark. Note that the charge of the W lepton and the soft electron from charm decay are opposite.	17

2.1	A schematic of the accelerator complex at Fermilab. CDF is located in the B0 collision hall, and DØ is located (appropriately) in the D0 collision hall.	20
2.2	Luminosity delivered and written to tape by store number. The data taking efficiency averages just over 80%.	23
2.3	Illustration CDF and its various subdetectors.	24
2.4	Cut-away illustration of the tracking volume in the x - y plane. Lines of constant pseudorapidity are labeled.	25
2.5	$r - \phi$ view of the L00, SVXII, and ISL subdetectors.	27
2.6	$r - \phi$ view of the L00 detector. Note the overlapping hexagonal structure.	28
2.7	Illustration of a cell in superlayer 2 of the COT.	29
2.8	Illustration of a calorimeter tower.	31
2.9	Illustration of a cross sectional view of the CMU, showing the path of a charged particle.	33
2.10	Central muon chamber coverage at CDF.	34
2.11	A block diagram of the CDF data acquisition system.	36
3.1	Fractional systematic uncertainty as a function of jet p_T . We do not use the ‘out-of-cone’ or ‘underlying event’ corrections.	47
3.2	An illustration of a primary and secondary vertex.	49
3.3	The tagging efficiency and mistag rate for loose and tight SECVTX as a function of the jet E_T	50
4.1	p_T spectrum of electrons from b hadron decay, c hadron decay, and photon conversion in $t\bar{t}$ MC events (generated with $M_{top} = 175 \text{ GeV}/c^2$). Distributions are normalized to unit area to show the shape.	53
4.2	qd'_0 for electrons from conversions, electrons from heavy flavor, and pions in $t\bar{t}$ MC events (generated with $M_{top} = 175 \text{ GeV}/c^2$). Distributions are normalized to unit area to show the shape.	54
4.3	An illustration of HF electrons, conversion electrons, and fake electrons embedded in jets.	55
4.4	Distributions of the selection requirements for the 8 GeV trigger electron. Shown are E_T (a), p_T (b), E_{EM}/p (c), E_{Had}/E_{EM} (d), L_{shr} (e), χ^2_{strip} (f), Δ_{wire} (g), and Δ_{strip} (h). The distributions are calculated before any requirement on a partner leg is made.	58
4.5	Electromagnetic fraction of jets matched to the conversion electron partner leg.	60
4.6	Illustration of conversion pairs with positive and negative R_{conv} and positive and negative sep	61

- 4.7 Conversion identification distributions, $\Delta \cot(\theta)$ (a), separation (b), and R_{conv} (c) in the 8 GeV electron dataset. Requirements are released on the given variables only to show the tails. $\Delta \cot(\theta)$ is fit to the sum of two Gaussians and a quadratic. The fit is shown in the solid line, while the quadratic component (which we associate with the combinatorial backgrounds) is shown in the dotted line. R_{conv} of the conversion electron pair with (solid line) and without (dotted line) the impact parameter and missing silicon requirements. The high R_{conv} conversions are suppressed due to the impact parameter requirement. The low R_{conv} conversions are suppressed due to the missing silicon layer requirement. 62
- 4.8 Track parameters for the conversion electron (solid line) and generic track (dotted line) samples. The scale and statistics for each are on the left and right, respectively. Parameters shown are p_T (a), qd'_0 (b), η_{trk} (c), and z_0 (d). 66
- 4.9 The isolation variable I distribution for the sample of conversion electrons and generic tracks. The embedded histogram shows the $I < 1.4$ region. The histograms have been normalized to unit area. As expected, conversion electrons are more isolated than generic tracks. 68
- 4.10 E_{EM}/p of electrons from the conversion sample and generic tracks normalized to unit area. Left: Isolated ($I = 1.0$). Right: Non-Isolated ($I > 1.0$). No E_{Had}/E_{EM} requirement is made. 69
- 4.11 E_{Had}/E_{EM} of electrons from the conversion sample and generic tracks normalized to unit area. Left: Isolated ($I = 1.0$). Right: Non-Isolated ($I > 1.0$). The embedded histograms show the $E_{Had}/E_{EM} < 0.22$ region. No E_{EM}/E_p requirement is made. 69
- 4.12 E_{EM}/p (top) and E_{Had}/E_{EM} (bottom) of the calorimeter clusters of electrons from b decay and cascade decay and fake electrons in $t\bar{t}$ MC events. 71
- 4.13 Number of wires in a CES cluster as a function of p_T (first two rows) and η (last two rows) for electrons (solid line) and tracks from jets (dotted line). Histograms have been normalized to unit area. 73
- 4.14 Number of strips in a CES cluster as a function of p_T (first two rows) and η (last two rows) for electrons (solid line) and tracks from jets (dotted line). E_{EM}/p and E_{Had}/E_{EM} requirements have been applied. Histograms have been normalized to unit area. 74
- 4.15 Combined efficiency of the $N_{wire} \geq 2$ and $N_{strip} \geq 2$ requirements on electrons and generic tracks as a function of p_T . Calorimeter requirements have been applied. Note the zero suppression. 75

- 4.16 χ^2 in wire (left) and strip (right) views, normalized to unit area. Electrons are shown in solid lines while generic tracks are shown in dotted lines. The histograms have been normalized to unit area, and the last bin is an overflow. Calorimeter and wire/strip requirements have been made. 77
- 4.17 χ_{wire}^2 for generic tracks. Calorimeter requirements only have been applied. Rows are divided between having the wire/strip requirement applied (bottom) and not applied (top). Columns are divided between p_T less than (left) and greater than (right) 5 GeV/ c . The second peak disappears when the 2 or more wires and 2 or more strips are required. The χ_{strip}^2 distributions are similar. 78
- 4.18 χ_{wire}^2 for conversion electrons (left) and generic tracks (right) in the wire view as a function of p_T . Electrons exhibit stable behavior over a broad range of transverse momentum, whereas generic tracks exhibit a longer tail at high p_T . All histograms have been normalized to unit area to show the relative shape. Calorimeter and wire/strip requirements have been made. 79
- 4.19 χ_{strip}^2 for conversion electrons (left) and generic tracks (right) in the wire view as a function of p_T . Electrons exhibit stable behavior over a broad range of transverse momentum, whereas generic tracks exhibit a longer tail at high p_T . All histograms have been normalized to unit area to show the relative shape. Calorimeter and wire/strip requirements have been made. 79
- 4.20 Mean χ^2 in the wire (left) and strip (right) view for conversion electrons (black) and generic tracks (gray) as a function of isolation. Calorimeter and wire/strip requirements have been made. 80
- 4.21 $q\Delta_{wire}$ as a function of p_T fit to the sum of two Gaussians constrained to have the same mean, μ . Calorimeter and wire/strip requirements have been made. 81
- 4.22 $q\Delta_{strip}$ as a function of p_T fit to the sum of two Gaussians constrained to have the same mean, μ . Calorimeter and wire/strip requirements have been made. 82
- 4.23 Electron $\langle q\Delta_{wire}/\sigma_{wire} \rangle$ (top) and $\langle \Delta_{strip}/\sigma_{strip} \rangle$ (bottom) as a function of p_T (left) and η (right). The errors shown are given by the RMS divided by \sqrt{N} , where N is the number of points. Thus, the large errors at high p_T correspond to fewer data-points, not larger RMS. Calorimeter and wire/strip requirements have been made. . . 83
- 4.24 Electron (left) and generic tracks (right) distributions of $q\Delta_{wire}/\sigma_{wire}$ (top) and $\Delta_{strip}/\sigma_{strip}$ (bottom) as a function of p_T . Each histogram has been normalized to unity to show the relative shape. Calorimeter and wire/strip requirements have been applied to both the electron and generic track samples. 84

4.25	Efficiency of a requirement made at that place on the likelihood. This is not the overall efficiency of the tagger since calorimeter and wire/strip requirements have already been applied.	85
4.26	Tagging efficiency as a function of p_T for conversion electrons (top) and generic tracks (bottom). Shown is the efficiency of the individual requirements on the left, and for each tag level on the right.	90
4.27	Tagging efficiency as a function of track η for conversion electrons (top) and generic tracks (bottom). Shown is the efficiency of the individual requirements on the left, and for each tag level on the right.	91
4.28	Tagging efficiency as a function of isolation for conversion electrons (top) and generic tracks (bottom). Shown is the efficiency of the individual requirements on the left, and for each tag level on the right.	92
4.29	Comparison of the isolation distribution for the generic track sample before and after the E_{EM}/p and E_{Had}/E_{EM} requirements.	93
5.1	Analytic calculation of conversion asymmetry distribution for a photon with P_T of 30 GeV/c and material with atomic number, $Z=13$. The distribution represents the differential probability that the positron will receive a given fraction of the photon's energy. Image taken from [1].	97
5.2	Radius of conversion for real conversions (top) and prompts modeled by anti-electrons (bottom). Conversion radius for anti-electrons can only be determined if another track is mistakenly paired with the original track. Conversions at high R_{conv} , such as at the COT inner cylinder, are suppressed due to the d'_0 cut.	103
5.3	Missing silicon layers versus the radius of conversion for a sample of conversions. The SVX detector ends at approximately 12 cm. The suppression of conversions at high R_{conv} is due to the tight impact parameter requirement.	104
5.4	Missing silicon layers versus expected silicon layers for conversion electrons (left) and anti-electrons (right) in the inclusive electron dataset. The distribution of expected silicon layers is dependent on the geometry of the track.	105
5.5	Over-efficiency as a function of the electron track p_T , η , z_0 , isolation, the number of tracks in the event, and the number of z vertices measured in events from the inclusive electron dataset.	108
5.6	Fit result for missing silicon layer templates.	111
5.7	Efficiency as a function of the electron track p_T , η , z_0 , isolation, the number of tracks in the event, and the number of z vertices measured in events from the inclusive electron trigger.	115
5.8	Data-MC Efficiency Scale Factor in the generic jet datasets as a function of p_T	119

5.9	Data-MC Efficiency Scale Factor in the generic jet datasets as a function of z_0	120
5.10	Data-MC Efficiency Scale Factor in the generic jet datasets as a function of the number of tracks in the event.	121
5.11	Data-MC Over-Efficiency Scale Factor in the generic jet datasets as a function of p_T	122
6.1	Invariant mass of high p_T trigger electron and candidate SLT_e track. The candidate track must have opposite sign, $p_T > 20$ GeV/ c , and pass taggable requirements.	127
6.2	Predicted versus Measured Level 1 SLT_e Tags in the conversion electron sample.	130
6.3	Predicted versus Measured Level 2 SLT_e Tags in the conversion electron sample.	131
6.4	Predicted versus Measured Level 3 SLT_e Tags in the conversion electron sample.	132
6.5	Combined E_{EM}/p and E_{Had}/E_{EM} efficiency for electrons from conversion in jet50 data (red) and MC (blue) as a function of p_T . Isolated tracks (top) are considered separately from non-isolated tracks (bottom). The difference between data and MC is also shown fitted to a constant.	134
7.1	E_{frac} distribution for taggable tracks in the Jet 50 data sample (left), $t\bar{t}$ MC sample (right). The distribution on the right is divided into electrons and non-electrons.	137
7.2	Fake electron tagging rate in different generic jet datasets as a function of p_T . SLT_e Tag Level 3 is used. No E_{frac} requirement is applied on the left, and $E_{frac} < 0.25$ is required on the right.	138
7.3	The effect of the $E_{frac} < 0.25$ requirement on the p_T distribution of taggable tracks in each of the four generic jet samples.	139
7.4	The E_{frac} distribution for taggable generic tracks with $10 < p_T$ (GeV/ c) < 12 in jet20 (left) and jet100 (right) MC. The fraction of electrons in the jet20 sample is 67% higher.	140
7.5	Tagging efficiency for generic tracks in the Jet 50 dataset. The lines indicate the multiplicative correction used in the fake matrix for the $ \eta > 0.8$ region.	140
7.6	Comparison of the isolation (a), p_T (b), and $\sum_{clst} p_T$ (c) distribution of taggable tracks in jet50 data and MC.	142
7.7	Uncorrected SLT_e L1 Fake Matrix measurement and prediction in the photon+jet sample as a function of the jet E_T	143
7.8	Uncorrected SLT_e L1 Fake Matrix measurement and prediction in the generic jet samples as a function of the track p_T and η	145

7.9	Uncorrected SLT _e L1 Fake Matrix measurement and prediction in the generic jet samples as a function of the track isolation and z_0	146
7.10	Uncorrected SLT _e L1 Fake Matrix measurement and prediction in the generic jet samples as a function of the jet E_T and the number of z vertices.	147
7.11	Uncorrected SLT _e L2 Fake Matrix measurement and prediction in the generic jet samples as a function of the track p_T and η	148
7.12	Uncorrected SLT _e L2 Fake Matrix measurement and prediction in the generic jet samples as a function of the track isolation and z_0	149
7.13	Uncorrected SLT _e L2 Fake Matrix measurement and prediction in the generic jet samples as a function of the jet E_T and the number of z vertices.	150
7.14	Uncorrected SLT _e L3 Fake Matrix measurement and prediction in the generic jet samples as a function of the track p_T and η	151
7.15	Uncorrected SLT _e L3 Fake Matrix measurement and prediction in the generic jet samples as a function of the track isolation and z_0	152
7.16	Uncorrected SLT _e L3 Fake Matrix measurement and prediction in the generic jet samples as a function of the jet E_T and the number of z vertices.	153
7.17	Weighted average of F_e^{tag} for tag level 1 combining generic jet datasets as a function of track p_T , η , and isolation bin in the fake matrix. Shown also is the average value with uncertainties.	162
7.18	On the left is the L_{xy} significance of track pairs requiring $0.4875 < M_{invariant} \text{ (GeV}/c^2) < 0.5075$. Embedded is the same plot with a logarithmic scale. On the right is the invariant mass of tracks pairs requiring L_{xy} significance > 8 . Track pairs are collected in the generic jet datasets.	164
7.19	These need to be remade. (Predicted-Measured)/Predicted tags in a sample of pions from K_s as a function of p_T , η , isolation, and E_{frac} . SLT _e tag level 1 is used.	165
8.1	Illustration of the $b\bar{b}$ event selection. We look for a Loose SECVTX tag and an 8 GeV lepton in the lepton jet. We look for a Tight SECVTX tag in the away jet. The SLT _e tagging efficiency is measured in the away jet.	169
8.2	Comparison of the beamline corrected impact parameter of taggable SLT _e tracks in the away jet of the $b\bar{b}$ sample. $b\bar{b}$ events are collected in the 8 GeV lepton samples.	170
8.3	Measured and predicted SLT _e level 1 tags in a $b\bar{b}$ sample as a function of the track p_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	176

8.4	Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the isolation. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	177
8.5	Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the beamline corrected impact parameter. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	178
8.6	Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the invariant mass between the trigger lepton and the SLT_e tag. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	179
8.7	Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the jet E_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	180
8.8	Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the relative p_T between the track and the jet. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	181
8.9	Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the track p_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	182
8.10	Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the isolation. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	183
8.11	Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the beamline corrected impact parameter. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	184
8.12	Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the invariant mass between the trigger lepton and the SLT_e tag. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	185
8.13	Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the jet E_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	186

8.14	Measured and predicted SLT _e level 2 tags in a $b\bar{b}$ sample as a function of the relative p_T between the track and the jet. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	187
8.15	Measured and predicted SLT _e level 3 tags in a $b\bar{b}$ sample as a function of the track p_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	188
8.16	Measured and predicted SLT _e level 3 tags in a $b\bar{b}$ sample as a function of the isolation. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	189
8.17	Measured and predicted SLT _e level 3 tags in a $b\bar{b}$ sample as a function of the beamline corrected impact parameter. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	190
8.18	Measured and predicted SLT _e level 3 tags in a $b\bar{b}$ sample as a function of the invariant mass between the trigger lepton and the SLT _e tag. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	191
8.19	Measured and predicted SLT _e level 3 tags in a $b\bar{b}$ sample as a function of the jet E_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	192
8.20	Measured and predicted SLT _e level 3 tags in a $b\bar{b}$ sample as a function of the relative p_T between the track and the jet. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.	193
9.1	Jet multiplicity of measured and Alpgen predicted $Z \rightarrow \ell^+\ell^-$ events at CDF ($\ell = e, \mu$). The MC is normalized to the data, which is dominated by the 0 jet bin. On the left is a comparison of the distributions, and on the right is the ratio.	204
9.2	\cancel{E}_T fits showing the fraction of QCD in the pretag (left) and tag (right) regions of the 1 jet bin of the lepton+jets sample.	205
9.3	\cancel{E}_T fits showing the fraction of QCD in the pretag (left) and tag (right) regions of the 2 jet bin of the lepton+jets sample.	206
9.4	\cancel{E}_T fits showing the fraction of QCD in the pretag (left) and tag (right) regions of the 3 jet bin of the lepton+jets sample.	207

9.5	Feynman diagram of $W + b\bar{b}$ event. This can be generated as a $W + b\bar{b}$ event by Alpgen or a $W + 1p$ event by Alpgen with a subsequent gluon split by Pythia.	213
9.6	Jet multiplicity of SLT_e tagged events in the lepton+jets dataset. The embedded plot is the ≥ 3 jet subsample.	215
9.7	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track p_T . tag track.	222
9.8	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track η . tag track.	222
9.9	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track z_0 . tag track.	223
9.10	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track d'_0 . tag track.	223
9.11	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track isolation. tag track.	224
9.12	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the invariant mass between the SLT_e tag and the tight W lepton. tag track.	224
9.13	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the event \cancel{E}_T . tag track.	225
9.14	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the event H_T . tag track.	225
9.15	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the reconstructed transverse mass of the W . tag track.	226
9.16	Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the lepton p_T/E_T . tag track.	226

List of Tables

1.1	Table of leptons and quarks in the Standard Model. Masses shown are the world average from the PDG [2].	3
4.1	Selection requirements for the 8 GeV trigger electron and primary leg of the conversion. See Section 3.2.1 for the definition of these variables.	57
4.2	Electron candidate selection requirements on conversion partner leg to the trigger.	59
4.3	Electron candidate selection requirements on conversion partner leg to the trigger.	61
4.4	The selection criteria for generic tracks.	65
4.5	Efficiencies of the $0.6 < E_{EM}/p < 2.5$ requirement and $E_{Had}/E_{EM} < 0.2$ requirement on different samples. Statistical errors are $< 1\%$ absolute.	72
4.6	The selection criteria to pass taggable requirements.	87
4.7	The selection criteria to pass tag requirements.	87
4.8	The selection criteria for anti-electrons.	88
5.1	Final conversion filter applied to each candidate electron track in an event. Tracks are removed from consideration when either a partner track is found or too many silicon layers are missing.	98
5.2	Summary of the sample selection described in this section. These samples are taken from events in which the inclusive electron trigger or the generic jet triggers were fired.	102
5.3	Conversion finding efficiency by different number of expected silicon layers.	112
5.4	Cross check measuring the efficiency in the generic jet datasets and comparing them against the true efficiency.	117
5.5	Measurement of the Partner Matching efficiency in data and comparison against the true efficiency in the corresponding MC. The first uncertainty is statistical, the second is systematic.	118

5.6	Measurement of the Missing Silicon Layer efficiency in data and comparison against the true efficiency in the corresponding MC.	118
5.7	Measurement of the combined conversion filter efficiency in data and comparison against the true efficiency in the corresponding MC. . . .	118
6.1	Constants for the η corrections in the forward, $ \eta > 0.8$, regions. . . .	125
6.2	Z electron sample selection summary.	126
6.3	Measured and Predicted Tags in the Z electron sample for a given Z mass window. The average difference is taken as a correction. The systematic error is inflated to cover the difference.	128
6.4	Z electron sample selection summary.	133
7.1	Constants for the η corrections in the forward, $ \eta > 0.8$, regions of the fake matrix.	141
7.2	Constants for the isolation corrections of the fake matrix.	141
7.3	Fraction of SLT_e tags that originate from conversions given by dataset and SLT_e tag level. The uncertainties shown for the efficiency and F_γ^{tag} are statistical followed by systematic. The uncertainties shown for the over-efficiency, ε^o , are systematic only (the statistical uncertainties are negligible).	156
7.4	Fraction of SLT_e tags that originate from the semi-leptonic decay of HF across dataset and SLT_e tag level.	158
7.5	Fraction of SLT_e tags that originate from other sources of electrons across dataset and SLT_e tag level.	160
7.6	Fraction of SLT_e tags that are real electrons in the generic jet datasets.	161
7.7	Selection requirements for K_s in events triggered on a high E_T jet.	163
7.8	Predicted and Measured SLT_e tags in the K_s pion sample by tag level.	164
8.1	8 GeV trigger electron selection criteria for the $b\bar{b}$ sample construction.	168
8.2	Muon selection criteria. The $Isol$ variable is the standard calorimeter isolation, not the track-based isolation normally used for the SLT_e	169
8.3	Per jet tagging rate in data and MC, for the inclusive electron and muon datasets, and by SLT_e tag level.	172
8.4	Away jet tagging rate in data and MC with the new tag matrix correction applied.	174
9.1	Summary of the data run taking periods and their corresponding luminosity. The luminosity shown has a common 6% relative systematic uncertainty.	196
9.2	Pretag event yield collected from the high p_T lepton datasets. We require $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV.	196
9.3	Tag event yield collected from the high p_T lepton datasets. We require $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV.	197

9.4	Uncorrected $t\bar{t}$ acceptance. We have required $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV. Statistical uncertainties are shown in parentheses.	198
9.5	A summary of scale factors and efficiencies used to correct the acceptance. Systematic uncertainties are shown in parentheses.	199
9.6	Combined lepton ID Scale Factor by lepton type and data period. These are the combined results from Table 9.5. Systematic uncertainties are shown in parentheses.	200
9.7	Corrected $t\bar{t}$ acceptance. We have required $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV. Combined statistical and systematic uncertainties are shown in parentheses.	200
9.8	Pretag and tag $t\bar{t}$ expectation assuming a top cross section of 6.7 pb. Uncertainties shown include the uncertainties associated with SLT_e tagging as well as lepton ID SFs, and luminosity.	201
9.9	Cross section for various processes that contribute to the lepton+jets dataset.	202
9.10	Cross section times branching fraction ($\ell = e, \mu, \tau$) calculated by Alpgen for various Z or Drell-Yan processes. The branching fractions for each lepton are identical due to lepton universality. The cross sections shown are then scaled by 1.68 ± 0.28 to match the measured $Z + jets$ spectrum.	203
9.11	Lepton+jets pretag composition for MC derived backgrounds. Uncertainties shown include lepton ID systematics, luminosity, and cross section uncertainties.	204
9.12	Lepton+jets SLT_e Level 1 Tag composition for MC derived backgrounds. Uncertainties shown include lepton ID systematics, luminosity, and cross section uncertainties, as well as systematic uncertainties associated with the SLT_e tagger.	205
9.13	Summary of the fraction of tag and pretag events in the lepton+jets channel due to QCD. These values have been derived from fits of the \cancel{E}_T in each of the jet bins.	208
9.14	Cross section times branching fraction ($\ell = e, \mu, \tau$) calculated by Alpgen for various W processes. The branching fractions for each lepton are identical due to lepton universality.	211
9.15	Heavy Flavor Fractions for W+Jet events. Uncertainties are systematic only, related to the K factor and the Q^2 scale of the generated MC. All numbers are shown in units of %.	212
9.16	SLT_e Level 1 tagging efficiency for different classes of HF in W+Jet events. Uncertainties shown include all SLT_e tagging systematics. All numbers are shown in units of %.	212

9.17	Sample Composition with SLT_e tag level 1. Uncertainties include luminosity, lepton SFs, cross section uncertainties, tagger systematics, K-factor, and the QCD fit.	216
9.18	Sample Composition with SLT_e tag level 2. Uncertainties include luminosity, lepton SFs, cross section uncertainties, tagger systematics, K-factor, and the QCD fit.	217
9.19	Sample Composition with SLT_e tag level 3. Uncertainties include luminosity, lepton SFs, cross section uncertainties, tagger systematics, K-factor, and the QCD fit.	218
9.20	Comparison of measured and expected tags in the signal region (≥ 3 jets) of the lepton+jets channel. SLT_e Tag levels shown are exclusive and assume $\sigma_{t\bar{t}} = 8.0$ pb unless otherwise indicated.	220
9.21	Summary of systematic uncertainties on the $t\bar{t}$ cross section measurement. Uncertainties shown are relative.	221
9.22	List of events in the lepton+jets signal region with at least 1 tag. . .	227

Acknowledgments

I am deeply indebted to a large number of people who made this thesis possible. First, I want to thank my family, mom, dad, and Joshua. My mother was my first science teacher and encouraged me along this path in too many ways, from reading books to me to throwing birthday parties at science museums. My dad sacrificed for the family and insisted that I go to the best schools, even when the finances weren't always in agreement. Even Josh has been an encouragement, calling me when I don't call back as much. I love you all.

I am also incredibly grateful for my advisor, Melissa Franklin, who not only admitted me into Harvard, but took me on as her student. Her uncanny ability to ask the right question at the right time routinely floors me. She never accepted the slightest hint of BS, and has constantly pushed me to become a better physicist. Her genuine care - expressed in her own way - did not go unnoticed.

I also want to thank all of the CDF grad students, post docs, and professors from Harvard, that made work a pleasure. Daniel Sherman, Ayana Holloway, Ingyin Zaw, Sal Rappoccio, Robyn Madrak, Kevin Burkett, Alexei Ershov, and Professor Andy Foland. Sebastian Grinstein was always around to listen and be a terrific sounding board for my wacky ideas. Professor João Guimarães da Costa, his care and attention to detail, is someone I have learned a lot from.

I am grateful to Aseet Mukherjee who took me under his wing and was extraordinarily patient with someone for whom analog electronics is not his strong suit. Aseet never fails to help me realize how little I understand about anything. Bill Ashmanskas was equally patient with me, as we worked on the Booster damper. He was a source of encouragement and helped me appreciate hardware in a way I hadn't before.

My thanks also go to George Brandenburg who was the professor when I TF'd undergraduate particle physics. He was a great teacher to learn from and to emulate. Peter Lu convinced me as a freshmen at Princeton that this whole Computer Science thing was a waste of time and later convinced me that Harvard was where it was at. Dan Babich was a great roommate at Harvard. Pinnocchio's runs would not have been nearly as fun. Thanks also to Nima Arkani-Hamed, Gary Feldman, and Matthew Schwartz for being on my various committees, and asking great questions.

And finally, I especially want to thank my wife, Hamille. She supported me when I needed it most, prayed for me continually, and taught me how *not* to think like a physicist when the time called for it. I love you very much.

How blessed is the man who does not walk in the counsel of the wicked,
Nor stand in the path of sinners,
Nor sit in the seat of scoffers.
But his delight is in the Law of *YHWH*,
And in His Law, he meditates day and night.

Soli Deo Gloria

Chapter 1

Physics of The Standard Model and Beyond

The Standard Model of particle physics (SM) is the best description of constitution of the physical world to date. For the last thirty years, hundreds of measurements varying in energy scale, detection apparatus, and analysis technique have all confirmed the basic structure of the SM without major modification. Only the direct observation of neutrino oscillations [3] provides a wrinkle in the canonical framework.

Nevertheless, there is presently an expectation that the SM is incomplete. Fundamental questions about the structure of the Universe suggest solutions beyond the SM. For instance: What is the origin of mass? How do we account for the matter/anti-matter asymmetry in the Universe? Do the electroweak and strong forces unify at some higher energy scale? Why is gravity so weak? What is the matter content of the Universe?

We begin this chapter with a description and history of the SM. We then discuss

the top quark, its place within the standard model, its *raison d'être*, and its properties.

1.1 The Standard Model

The Standard Model is a Quantum Field Theory (QFT) consisting of three generations of quarks and three generations of leptons which interact with each other via vector bosons. Within each generation, the quarks and leptons appear in pairs, and each successive generation has identical quantum numbers, but different masses. Table 1.1 illustrates this generational structure and the corresponding mass hierarchy. The photon, W and Z bosons, and gluons are the intermediate vector bosons responsible for “carrying” the electromagnetic, weak, and strong forces, respectively. The gravitational force is not included in the SM since a QFT of gravity is necessarily non-renormalizable. Whether or not the Higgs boson is included in the SM is arguably a matter of taste. Although the Higgs mechanism is the leading explanation for electroweak symmetry breaking, it is not the only, and there is not yet any evidence for its existence.

The three generation structure of the SM was proposed in 1973 by Kobayashi and Maskawa [4]. Glashow, Iliopoulos, and Maiani (GIM) had suggested that the anomalously low $K^0 \rightarrow \mu^+ \mu^-$ rate could be explained by the introduction of a fourth quark [5], and Kobayashi and Maskawa extended the argument in an attempt to introduce CP violation directly into the scheme. But it was not until the discovery of the J/ψ [6, 7] in the November Revolution that the existence of the charm quark (let alone a third generation) became evident. The first direct evidence of the third generation soon came with the discovery of the τ lepton by Martin Perl [8]. Next, a resonance

Quarks		
up (u) $m_u \sim 1.5 - 3.0 \text{ MeV}/c^2$	charm (c) $m_c = 1.25 \pm 0.09 \text{ GeV}/c^2$	top (t) $m_t = 172.5 \pm 2.7 \text{ GeV}/c^2$
down (d) $m_d \sim 3 - 7 \text{ MeV}/c^2$	strange (s) $m_s = 95 \pm 0.25 \text{ MeV}/c^2$	bottom (b) $m_b = 4.20 \pm 0.07 \text{ GeV}/c^2$
Leptons		
electron (e) $m_e = 0.5110 \text{ MeV}/c^2$	muon (μ) $m_\mu = 105.6 \text{ MeV}/c^2$	tau (τ) $m_\tau = 1776.90 \pm 0.20 \text{ MeV}/c^2$
electron neutrino (ν_e)	muon neutrino (ν_μ)	tau neutrino (ν_τ)

Table 1.1: Table of leptons and quarks in the Standard Model. Masses shown are the world average from the PDG [2].

in the $\mu^+\mu^-$ invariant mass spectrum was discovered and quickly interpreted as a $b\bar{b}$ bound state [9]. The final pieces of the three generation structure of quarks and leptons were confirmed by the discovery of the top quark in 1995 [10, 11] and the tau neutrino, ν_τ , in 2001 [12]. From precision measurements of the Z boson width, the possibility of a 4th generation neutrino with mass $\lesssim M_Z/2$ is excluded [13]. This suggests that the 3rd generation of the SM may indeed be last, but the evidence is far from conclusive.

Is there any physics beyond the Standard Model? The abundance of unresolved questions seems to suggest so. Nevertheless, even indirect physical evidence is hard to come by. Searches for proton decay, sensitive to different Grand Unified Theory (GUT) models, continue to set limits on the lifetime of the proton [14]. Supersymmetric loop contributions should evince themselves in the anomalous magnetic moment of the muon, but the measured $(g - 2)_\mu$ is consistent with the SM at the $\sim 3\sigma$ level [15]. Similarly, the rare decay $b \rightarrow s\gamma$ is sensitive to extra-SM physics, but so far,

the experimental average $(3.54 \pm 0.26) \times 10^{-4}$ [2] is consistent with the SM prediction [16]. The answer to this question will likely be made known soon as the Large Hadron Collider at CERN turns on in the coming Spring.

1.2 The Top Quark

1.2.1 Expectation

In the intervening years between the discovery of bottom quark and the top quark, precision experiments examining the b quark's decay properties, electric charge, and weak isospin demonstrated that the b should be a part of a quark doublet. The existence of its isospin partner, the top quark, could then be inferred through its presence in the loops of various Feynman diagrams. For instance, the decay of B^0 and B_s to di-muons should be suppressed through the GIM mechanism if the b is part of an isospin doublet (see Figure 1.1). If the top quark did not exist, then the measured branching fractions would be much higher than observed (see [17] for the best current limit).

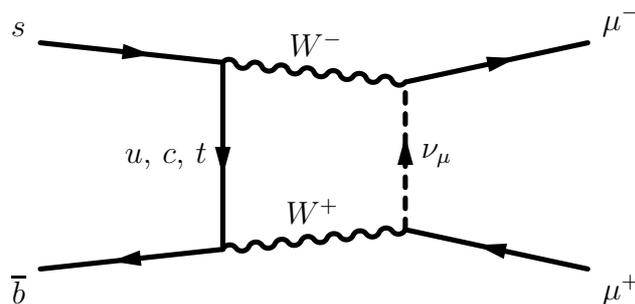


Figure 1.1: Feynman diagram showing the SM tree order process of $B_s \rightarrow \mu^+ \mu^-$.

It also turns out that completion of the doublet is necessary to preserve gauge in-

variance of the Glashow-Weinberg-Salam (GWS) electroweak theory. By forcing the weak interaction to couple only to left-handed particles (and right-handed antiparticles), gauge invariance can be broken at higher orders in perturbation theory. The Feynman diagram responsible for this effect, known as an anomaly, is shown in Figure 1.2, where the external lines are gauge bosons, and the internal lines are all possible fermion species. In theories like Quantum Electrodynamics (QED) or Quantum Chromodynamics (QCD) where the vector boson couples equally to both left- and right-handed species, anomalous diagrams are automatically canceled.

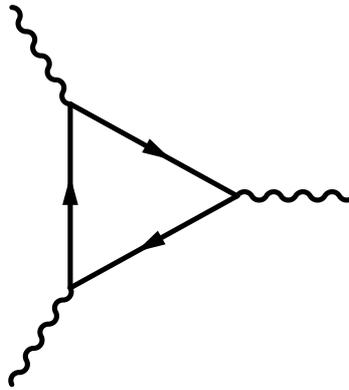


Figure 1.2: Triangle diagram, showing the anomalous term. The diagrams needed to show that the electroweak theory is free of anomalies has three external gauge bosons, and the internal lines are all possible fermions.

It is sufficient to show that the anomaly disappears in the full quantum theory for electroweak interactions if the trace over all possible fermions of the triangle diagram with three external gauge bosons cancels (more diagrams need consideration for a full theory involving QCD and gravity). This is true only if bottom has an isospin partner and if the number of generations of fermions and quarks is equal.

1.2.2 Mass

Although there was ample evidence for the existence of the top quark for many years, when the discovery was finally announced by the CDF and D0 collaborations in 1995, the surprise was its mass. The top quark is the most massive particle in the SM, and, for the time being, can only be produced in a laboratory setting¹ at the Tevatron at Fermilab. Its high mass relative to the other SM particles is a curiosity and results in a unique set of properties. For example, the high top mass results in lifetime shorter than Λ_{QCD} . This means that top decays before hadronization can occur, and measurements of top probe a bare quark. Measuring the mass of the top quark is also an interesting platform for probing new physics.

One of the most important contributions of the top quark mass to the understanding of new physics is through radiative corrections to the W mass. At tree level, the mass of the W can be related back to the Fermi constant, G_F , and the Z mass (via the Weinberg angle) according to the equation

$$M_W^2 \left(1 - \frac{M_W^2}{M_Z^2} \right) = \frac{\pi\alpha}{\sqrt{2}G_F} \quad (1.1)$$

where α is the fine structure constant. Radiative corrections to the W mass can be incorporated by rewriting Equation 1.1 as

$$M_W^2 \left(1 - \frac{M_W^2}{M_Z^2} \right) = \frac{\pi\alpha}{\sqrt{2}G_F} \left(\frac{1}{1 - \Delta r} \right) \quad (1.2)$$

where

$$\Delta r = \Delta\alpha + \Delta\rho(M_{top}^2) + \Delta\chi(\ln(M_H/M_Z)), \quad (1.3)$$

¹Top quarks should be produced through ultra high energy cosmic ray collisions against the Earth's atmosphere, but observation of this effect has not been achieved.

which includes the running of α , and the quadratic and logarithmic dependencies on the mass of the top and Higgs, respectively. The Feynman diagrams responsible for the latter two corrections are shown in Figure 1.3.

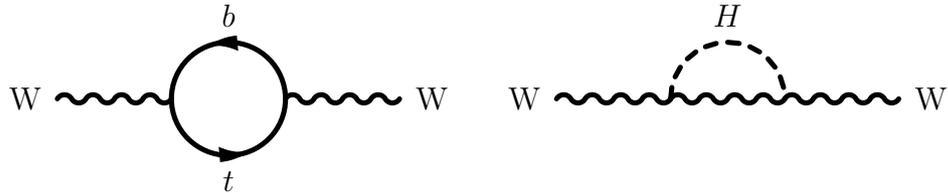


Figure 1.3: Feynman diagram showing radiative corrections from the top quark and Higgs to the W mass.

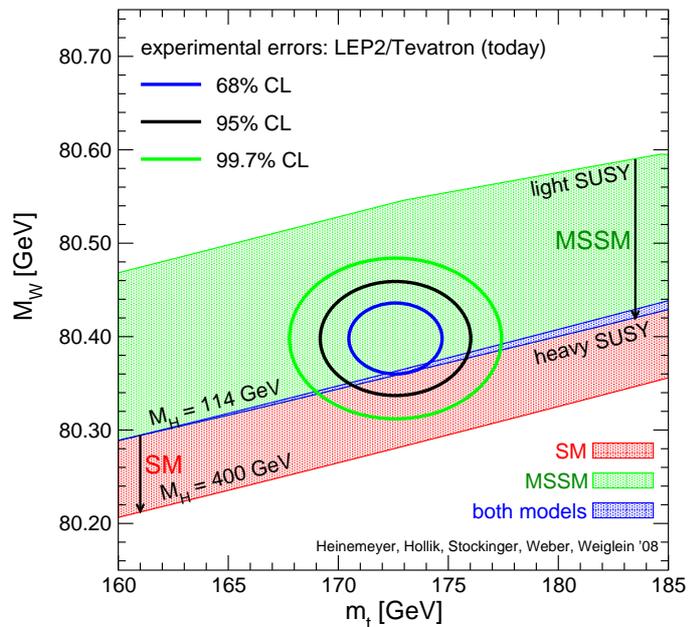


Figure 1.4: Higgs mass expected from precision measurements of the top and W mass. Shown are the theoretical regions favored by a purely SM Higgs and a Higgs in the MSSM. Assumed values: $M_{top} = 172.6 \pm 1.4 \text{ GeV}/c^2$ and $M_W = 80.398 \pm 0.025 \text{ GeV}/c^2$.

By making precision measurements of α , G_F , M_Z , M_W , and M_{top} , we can infer the mass of the Higgs. Figure 1.4 shows the Higgs mass as a function of M_W and M_{top} .

Higgs masses preferred by the SM and by the Minimal Supersymmetric Standard Model (MSSM) are both shown [18]. Using direct search through e^+e^- annihilation at LEP, a neutral Higgs has been ruled out with mass below $114.4 \text{ GeV}/c^2$ [19].

1.2.3 Hadroproduction

Pair Production

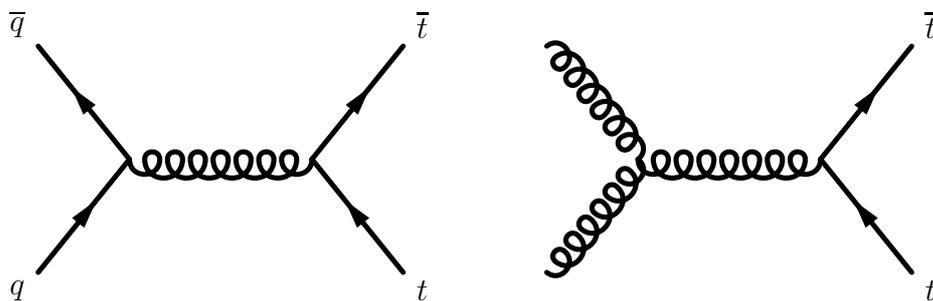


Figure 1.5: Feynman diagrams of quark annihilation (left) and gluon fusion (right). These are the two prominent production mechanisms of top at the Tevatron.

In $p\bar{p}$ collisions, top is typically pair produced via quark annihilation or gluon fusion. Feynman diagrams of these processes are shown in Figure 1.5. At the Tevatron, using $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$, $t\bar{t}$ is produced approximately 85% through quark annihilation and 15% through gluon fusion. At the LHC using pp collisions at $\sqrt{s} = 14 \text{ TeV}$, these ratios will effectively reverse, and gluon fusion will dominate by a factor of 4. At leading order (LO) in perturbation theory, the cross section through quark annihilation is given by:

$$\hat{\sigma} = \frac{8\pi\alpha_s^2}{27\hat{s}} \sqrt{1 - \frac{4m_t^2}{\hat{s}}} \left(1 + \frac{2m_t^2}{\hat{s}}\right) \quad (1.4)$$

where $\sqrt{\hat{s}}$ is the center of energy of the $q\bar{q}$ system.

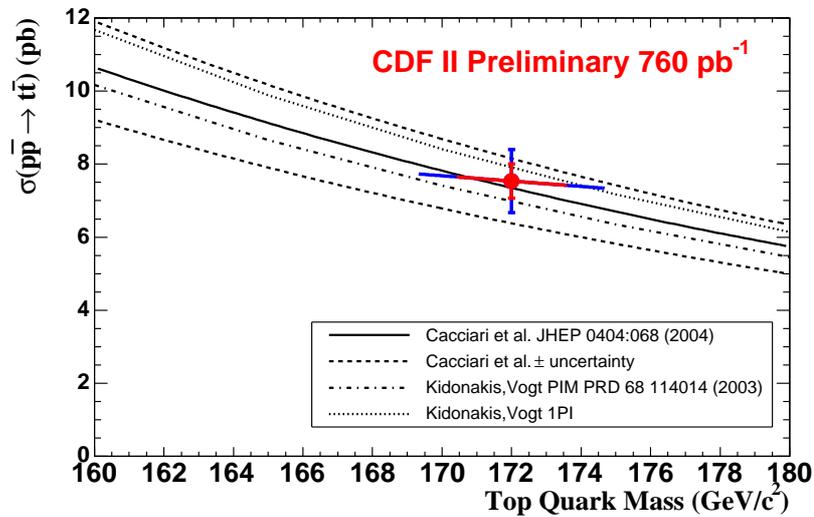


Figure 1.6: $t\bar{t}$ production cross section as a function of the top mass. Shown is the combined CDF Run 2 result with 760 pb^{-1} .

The current theoretical calculation of the production cross section at the Tevatron yields an expected value of $\sigma_{t\bar{t}} = 6.7 \pm 0.8 \text{ pb}$ depending on the method implemented [20, 21]. This value is approximately 10 orders of magnitude lower than the total $p\bar{p}$ inelastic cross section. The cross section has a dependence on both the assumed top mass (Figure 1.6) and the \sqrt{s} at which the top pair was produced (Figure 1.7). The dominant uncertainty comes from the renormalization and factorization scales (μ_R and μ_F , respectively) used in the perturbative calculation and from the parton distribution functions (PDFs) of the proton. We discuss each of these systematics briefly.

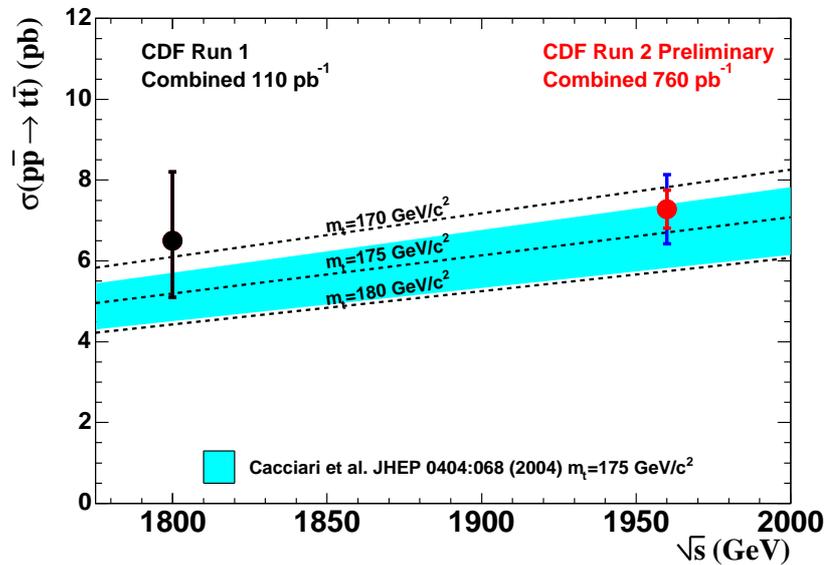


Figure 1.7: $t\bar{t}$ production cross section as a function of the \sqrt{s} energy. Shown are the CDF Run 1 and CDF Run 2 results.

Parton Distribution Functions

PDFs are the probability density associated with finding a particular parton (either quark or gluon) inside of the proton with a given fraction of the total momentum. Phenomenologically, the proton consists of massless partons within some hard boundary, beyond which confinement takes over [22]. Quantum fluctuations allow the presence of any number of types of partons, not just the up and down quarks.

The probability of finding a parton with a given momentum fraction depends on the momentum transfer scale, Q . Figure 1.8 shows the PDFs calculated by the Coordinated Theoretical Experimental Project on QCD (CTEQ) with updated information from QCD studies at HERA and the Tevatron [23] for $Q = 2$ GeV and 100 GeV. Uncertainties on the PDFs result from their basis in experimental mea-

measurements, and the uncertainty on the top cross section is derived from varying by $\pm 1\sigma$ the value of the fit parameter eigenvalues after diagonalization of the correlation matrix.

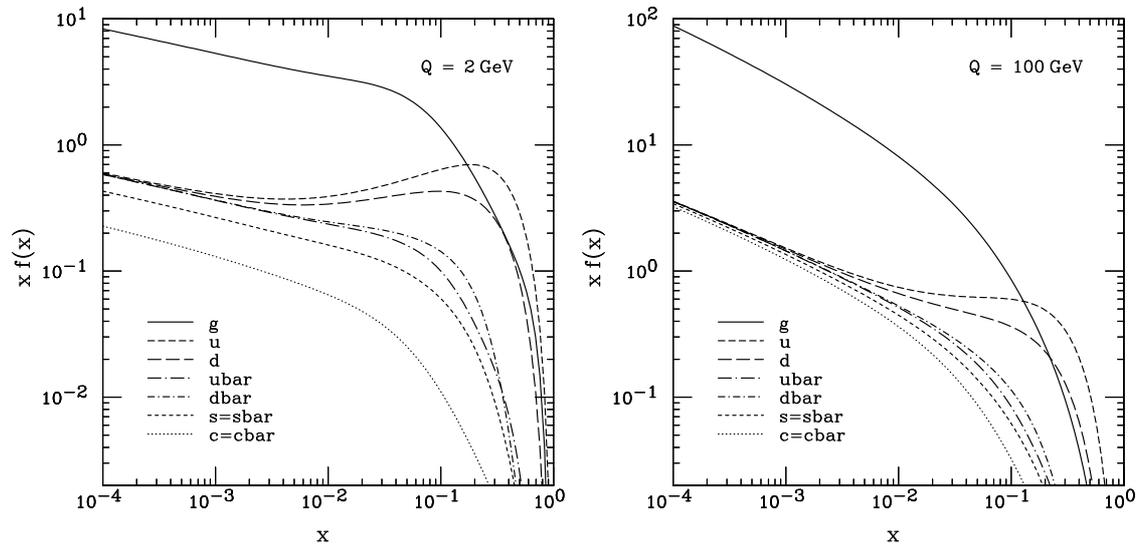


Figure 1.8: Next-to-leading order PDFs for protons from the CTEQ6M parameterization, where x is the momentum fraction, and $xf(x)$ is the probability density. $Q = 2$ and 100 GeV on the left and right, respectively.

At the Tevatron, top pair production requires a large fraction of the total energy of the $p\bar{p}$ collision. Since the up and down quarks carry the largest fraction of momentum in the proton, pair production will preferentially be through quark annihilation. On the other hand, at the LHC, top production will not require the same demands on the energy fraction, hence the preference for gluon fusion.

Renormalization/Factorization Scale

The renormalization and factorization scales are arbitrary scales chosen to allow for the calculation of various physical constants, such as the top production cross

section. In calculations involving Quantum Chromodynamics (QCD), the effective coupling, α_s has an energy scale dependence. The running of the coupling constant is determined by the β -function. Calculation of the cross section involves finding the coefficients in the term

$$\sigma = A_1\alpha_s + A_2\alpha_s^2 + \dots \quad (1.5)$$

from the appropriate Feynman diagram. Note that this series converges only when α_s is small (at high energy).

As is typical with QFT calculations, divergences arise in the calculation of the coefficients and must be regulated in a consistent way. In most regularization schemes, such as the modified minimal subtraction scheme (\overline{MS}), an arbitrary mass scale, μ_R , is introduced *ad hoc* to enforce dimensionally correct variables. Therefore, A_i depend explicitly on μ_R . In principle, σ should have no dependence on μ_R so long as every term in the series is properly calculated. However, in practice, calculating more than the first few terms is not possible, and a dependence is introduced.

Similarly, μ_F is an arbitrary scale where the calculation is separated into two factors, a perturbative, short-distance component and an non-perturbative, long-distance component (when α_s runs high). The short-distance phenomena is calculated using perturbative methods, but the long-distance phenomena is assumed to be part of the hadron structure and it is absorbed into a modified parton distribution function.

In practice, an appropriate choice for these scales should minimize the dependence. For the top cross section, the renormalization and factorization scales are set to m_{top} and varied from $m_{top}/2$ to $2m_{top}$ to determine the uncertainty associated with this

choice of scale. At NLO, this results in approximately a 10% fractional uncertainty on $\sigma_{t\bar{t}}$ [20].

Single Top Production

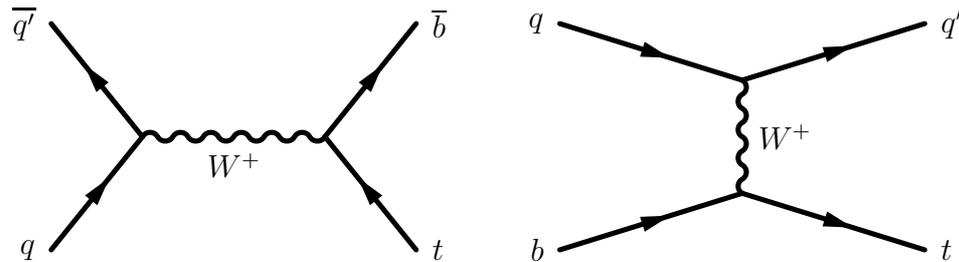


Figure 1.9: Feynman diagrams showing single top production in the s-channel (left) and t-channel (right).

Single top quarks can also be produced at the Tevatron through weak charged current (CC) channels. The primary tree order Feynman diagrams responsible are shown in Figure 1.9. At the Tevatron (again, using $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV), the theoretical cross section is $\sigma_{top} = 2.9 \pm 0.4$ pb assuming $m_{top} = 175$ GeV/ c^2 [24]. Although the tree level diagram is electroweak, next-to-leading order (NLO) corrections have gluon loops, so the calculation has all of the difficulties intrinsic to QCD calculations. The cross section was calculated with multiple regularization techniques, and the uncertainty is dominated by the factorization and renormalization scales, PDFs, and the mass of the b -quark. As of this year, evidence for single-top production has been seen at both CDF and D0 [25].

An interesting property of this measurement is that single-top production has a strong dependence on the Cabibbo-Kobayashi-Maskawa (CKM) matrix element $|V_{tb}|$. The CKM matrix determines the amount of mixing between the weak eigenstates and

mass eigenstates of the quark sector. A measurement of σ_{top} will be able to evaluate the element without assuming the three generation SM structure [26]. Assuming unitarity of the CKM matrix and three generations gives a value of $V_{tb} \simeq 1$ [2].

1.2.4 Decay

Assuming a SM value for V_{tb} of unity, top will decay exclusively into a W boson and b quark. Alternative decay channels such as $t \rightarrow Ws$ and $t \rightarrow Wd$ are suppressed by the factors $|V_{ts}|^2/|V_{tb}|^2 \approx 10^{-3}$ and $|V_{td}|^2/|V_{tb}|^2 \approx 10^{-4}$, respectively. Flavor changing neutral current (FCNC) decays, such as $t \rightarrow Zc$ and $t \rightarrow Z\gamma$, do not occur at tree level in the SM but through Penguin diagrams and are highly suppressed. Any observation would indicate physics beyond the SM [27].

For the purposes of reconstructing top, we can characterize the top pair decay modes into whether or not the subsequent W bosons decay leptonically (either e or μ), tauonically, or hadronically. The distinction between light and heavy leptons is pragmatic, reflecting the difficulty in reconstructing τ relative to the lighter leptons. Since W couples equally to quarks and leptons, the W decay branching fractions are roughly equal between all quark and lepton species after accounting for the three quark colors: $BF(W^- \rightarrow \ell\bar{\nu}_\ell) \approx 1/9$, where $\ell = e, \mu, \tau$, and $BF(W^- \rightarrow q\bar{q}') \approx 1/3$, where $q\bar{q}' = d\bar{u}$ and $s\bar{c}$.

The four canonical $t\bar{t}$ decay channels are:

- *all hadronic channel* ($BF \approx 44\%$): Both W 's decay hadronically. This decay channel has the highest branching fraction but also the highest level of irreducible backgrounds due principally to QCD production.

- *lepton+jets channel* ($BF \approx 30\%$): One W decays leptonically (e or μ) and the other hadronically. For many analyses, this channel offers the best tradeoff between signal purity and acceptance. For the analysis presented in this dissertation, we use this channel to measure $t\bar{t}$ events. The dominant background is a W boson produced in association with jets. Figure 1.10 shows a Feynman diagram of $t\bar{t}$ production and subsequent decay into this channel.
- *tauonic channel* ($BF \approx 21\%$): One W decays to a τ . This is a challenging decay channel because - as previously mentioned - τ is relatively difficult to reconstruct. Fake τ leptons dominate the background.
- *dilepton channel* ($BF \approx 5\%$): Both W 's decay to a light lepton. This channel offers the best signal purity at the cost of a signal acceptance. The dominant background to this measurement is Drell-Yan production of oppositely-charged dilepton pairs.

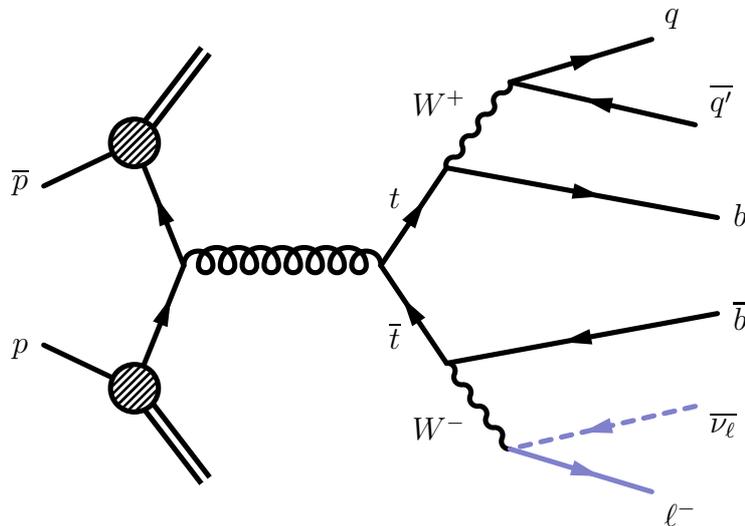


Figure 1.10: Feynman diagram of tree level $t\bar{t}$ production and subsequent lepton+jets decay. The lepton and neutrino have been highlighted to note the decay channel.

1.3 Soft Electron Tagging Measurements

1.3.1 $t\bar{t}$ Cross Section

In this dissertation, we measure the $t\bar{t}$ cross section using $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Top pair events are reconstructed in the lepton+jets channel, assuming $V_{tb} = 1$. In order to enhance the significance of the $t\bar{t}$ signal over the background, we tag b -jets using soft electrons from the decay of heavy flavor (HF), that is, b or c quarks. The branching fraction for B hadrons to decay semileptonically is approximately 10%. The subsequent decay of B hadrons to charm also has a semileptonic component: $BF(b \rightarrow c \rightarrow \ell\nu_\ell X) \approx 10\%$.

This measurement explores the properties of the top quark. Any significant deviation of the cross section measurement from either the theoretical value or the measured cross section in other channels could be an indication of new physics. At CDF in Run 1, for instance, there was measured an excess of events in the lepton+jets channel when b -jets were identified with a combined secondary vertex tagger and a soft lepton tagger [28]. Figure 1.11 shows the most recent results at CDF of the top production cross section, including the result described in this thesis.

1.3.2 Production of W plus Charm

This measurement also provide a platform for measuring the cross section of W in association with a single charm quark. At tree level, the electric charge of the W lepton is anti-correlated with the electric charge of the lepton from the semileptonic decay of charm (see Figure 1.12). Since many of the SM backgrounds have no charge

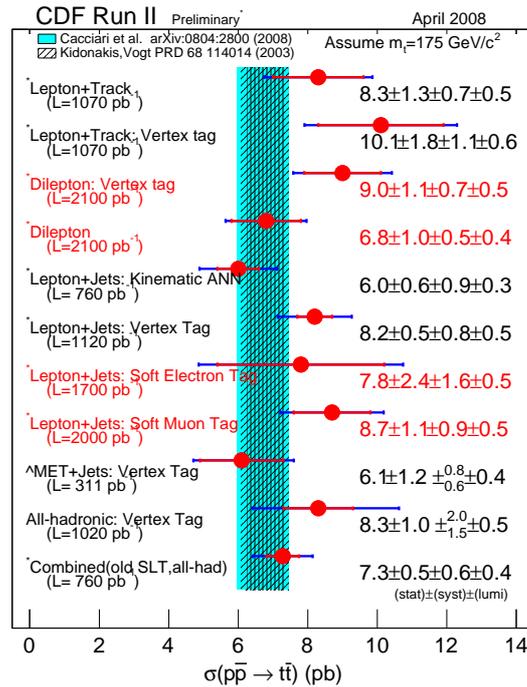


Figure 1.11: Combined results from CDF of the $t\bar{t}$ production cross section. Results shown include the cross section result from this measurement.

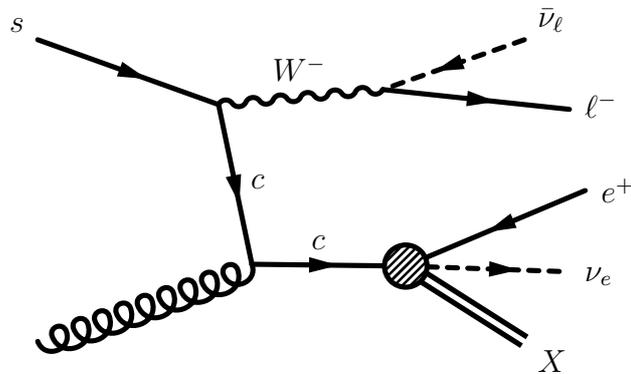


Figure 1.12: Tree order diagram of W hadroproduction associated with a single charm quark. Note that the charge of the W lepton and the soft electron from charm decay are opposite.

correlation asymmetry, the W +charm cross section can be measured by subtracting the number of same-sign events from the number of opposite-sign events. This measurement is complementary to the top cross section measurement, because the event

signature is the same except for the number of expected jets.

Chapter 2

Experimental Apparatus

The Tevatron refers to last stage of an entire accelerator chain located at Fermilab in Batavia, IL. Protons and antiprotons are accelerated to an energy of 960 GeV in opposite directions to make collisions with a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. The current physics run, called Run 2, began in 2002 and is ongoing. Run 1 occurred from 1992-1996, and had $\sqrt{s} = 1.8$ TeV.

Collisions occur at two points around the Tevatron, where two multipurpose detectors, CDF and DØ, reconstruct the outcome. In this chapter, we will discuss the accelerator complex and CDF, the detector used in our measurement.

2.1 The Accelerator Complex

2.1.1 Proton Acceleration: From Source to Main Injector

The accelerator complex begins with pure hydrogen gas which is ionized to H^- . The ions are accelerated from rest to 750 keV by a Crookroft-Walton electrostatic

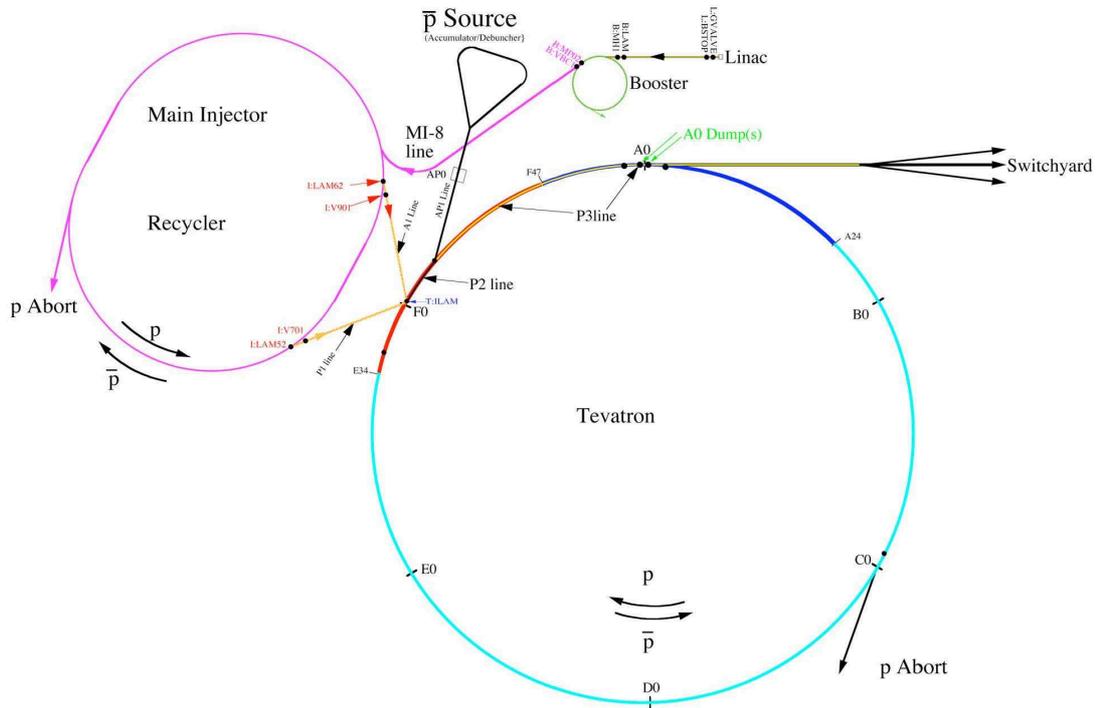


Figure 2.1: A schematic of the accelerator complex at Fermilab. CDF is located in the B0 collision hall, and $D\bar{0}$ is located (appropriately) in the D0 collision hall.

generator. The ions are steered and focused by magnets down a transfer line to a linear Radio Frequency (RF) accelerator, the Linac. The Linac further accelerates the ions to 400 MeV, at which point they become bunched together.

From the Linac, ions are transferred to a 15 Hz synchrotron, called the Booster. Here the ions hit a target, stripping the electrons from the proton. The Booster accelerates the protons from 400 MeV to 8 GeV around a 75-meter radius circle using 18 RF cavities interspersed throughout the circle. Combined function magnets bend and focus the beam.

The protons are sent from the Booster to the Main Injector where they are accelerated to 150 GeV for Tevatron injection or 120 GeV for antiproton production.

Protons can also be sent to Neutrinos at the Main Injector (NuMI) beamline or other fixed target experiments. The Main Injector is a synchrotron consisting of 344 dipole magnets and 208 quadrupole magnets for steering and focusing, respectively.

2.1.2 Antiproton Production

The number of $p\bar{p}$ collisions at the Tevatron is limited primarily by the ability of the accelerator to produce and store antiprotons. Antiprotons are produced by colliding 120 GeV protons from the Main Injector against a nickel target. Approximately 1 antiproton with an average momentum of 8 GeV/ c is produced for every 100,000 protons on target. The resulting antiprotons are sent down the AP2 transfer line to the Debuncher, a triangular synchrotron used to capture antiprotons.

When another batch of antiprotons arrives, the antiprotons presently in the Debuncher are sent to the Accumulator, another triangular synchrotron, housed in the same tunnel as the Debuncher. The Accumulator stores the antiprotons and cools them with a variety of techniques, such as stochastic cooling. In the Accumulator, antiprotons can be stored for days with minimal losses.

When the Accumulator is filled to capacity, antiprotons are sent to the Recycler, a permanent magnet storage ring inside the same tunnel as the Main Injector. The Recycler keeps the antiprotons at 8 GeV and further cools them with a technique called electron cooling. Antiprotons can then be sent to the Main Injector for acceleration up to 150 GeV and then injection into the Tevatron.

2.1.3 Tevatron

The Tevatron is a 1 km radius synchrotron accelerator capable of accelerating protons and antiprotons in the same beam pipe. The beam is steered with 774 superconducting dipole magnets and focused with 240 quadrupole magnets. The niobium-titanium magnets are cooled to 4.2K at which point they become superconducting. The advantage of superconducting magnets is that they admit higher currents without damage due to resistive heating.

For physics runs, 36 bunches of protons and 36 bunches of antiprotons are injected into the Tevatron. This results in a crossing time of 396 ns. Once the beam is accelerated, the protons and antiprotons may continue for hours (or even days) until either a mechanical failure or the number of collisions is deemed too low to be useful. The number of collisions is described by the instantaneous luminosity and is given by the formula:

$$\mathcal{L} = \frac{N_p N_{\bar{p}} N_B f}{2\pi \sigma_p^2 \sigma_{\bar{p}}^2} \quad (2.1)$$

where N_B is the the number of bunches of N_p protons and $N_{\bar{p}}$ antiprotons colliding at a frequency f . The values σ_p and $\sigma_{\bar{p}}$ are the effective widths of those bunches. The integrated luminosity, $\int \mathcal{L} dt$, gives a measure of the number of collisions in a fixed amount of time. If a physical process $p\bar{p} \rightarrow X$ has a cross section, σ , then the expected number of events is $N = \sigma \int \mathcal{L} dt$. In the analysis presented, we use an integrated luminosity of $\int \mathcal{L} dt = 1.7 \text{ fb}^{-1}$, corresponding to data taken up to March of 2007. By comparison, the total integrated luminosity for Run I was 126 pb^{-1} . Figure 2.2 shows the integrated luminosity delivered to CDF and stored to tape. The data taking efficiency averages just over 80%.

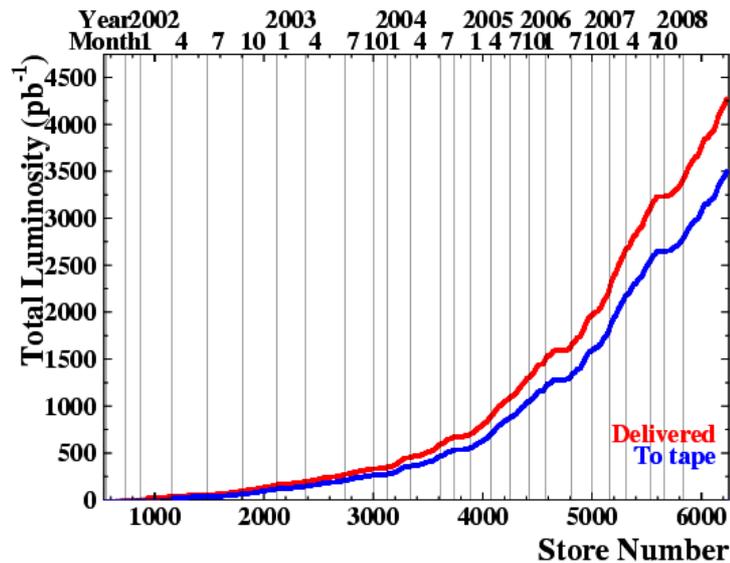


Figure 2.2: Luminosity delivered and written to tape by store number. The data taking efficiency averages just over 80%.

2.2 The Collider Detector at Fermilab

CDF is a multi-purpose detector described in [29, 30] and designed to study $p\bar{p}$ collisions at the Tevatron. The detector uses a combination of different tracking chambers, electromagnetic and hadronic calorimetry, and muon chambers to suit various types of physics analyses from precision B physics to searches for extra-SM particles.

The CDF coordinate system is right-handed with the z -axis pointing along the direction of the protons. The x - and y -axes point outward from the Tevatron ring and upward, respectively. It is also convenient to define the azimuthal angle, $\phi \equiv \tan^{-1}(y/x)$, and polar angle, $\theta \equiv \cos^{-1}(z/\sqrt{x^2 + y^2 + z^2})$. We can also define the pseudorapidity, $\eta \equiv -\ln(\tan(\theta/2))$. Often, pseudorapidity is used instead of θ because $\Delta\eta$ is Lorentz invariant in the massless approximation. This is especially rel-

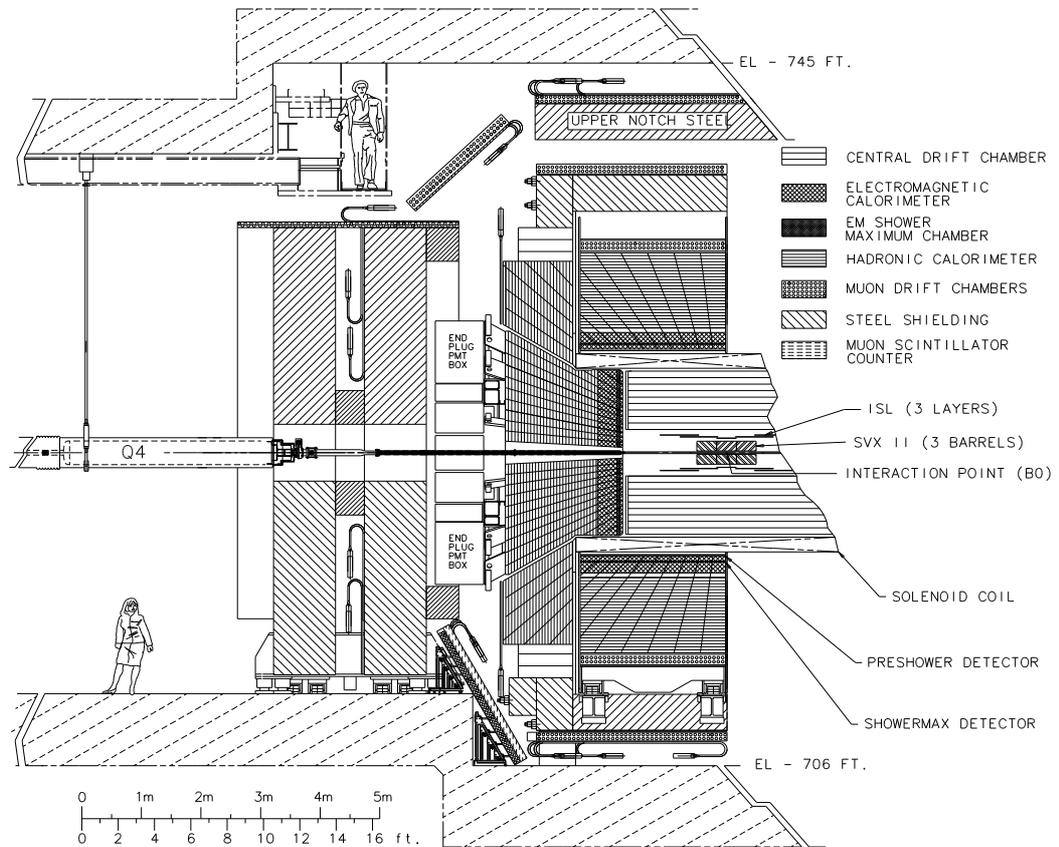


Figure 2.3: Illustration CDF and its various subdetectors.

event when measuring the number of particles per unit rapidity, $dN/d\eta$, which is invariant under boosts in the z direction. Figure 2.3 shows a schematic illustration of CDF. Figure 2.4 shows an illustration of CDF projected in the x - y plane with constant pseudorapidity lines shown.

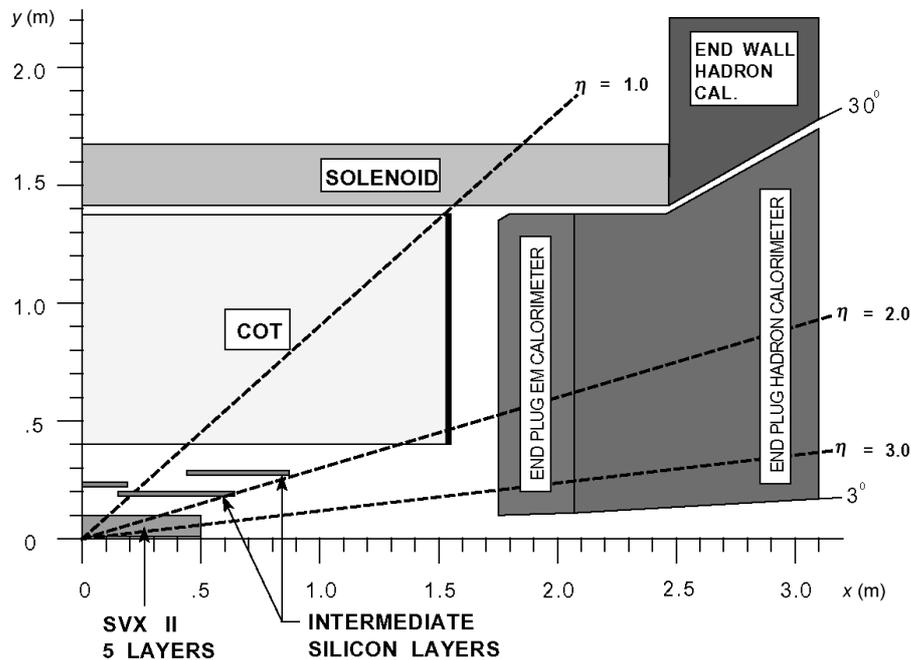


Figure 2.4: Cut-away illustration of the tracking volume in the x - y plane. Lines of constant pseudorapidity are labeled.

2.2.1 Tracking

The tracking system can be subdivided into two components, silicon vertex detectors and a drift chamber. The silicon vertex detectors lie close to the beam pipe and are intended to make precision measurements of track vertices. The primary application of the silicon detector is for identifying b -jets, which have a long lifetime relative to other jets. The silicon detectors surround the beam pipe and extend from 1.2 cm to 28 cm in radius. The cylindrical drift chamber that surrounds the silicon detector is less precise but extends from a radius of 44 cm to 132 cm.

The CDF tracking system is immersed in a uniform 1.4 Tesla magnetic field that runs parallel to the beamline and is generated by a niobium-titanium superconducting

solenoid magnet. The solenoid itself has a radius of 1.5 m, and is 5 m in length. The magnetic field allows for electric charge separation of tracks and a measurement of their momentum.

Charged particles that originate from the beamline will travel in a helical path due to the presence of the magnetic field. The five parameters which fully describe the path of a charged particle are:

- d_0 , the impact parameter. This is the distance of closest approach of the track to the location $(0, 0)$ inside the CDF detector. Often the impact parameter is corrected to be relative to the location of the beamline instead of $(0, 0)$.
- z_0 , the z -coordinate of the track at the point of closest approach.
- C , the half curvature defined as $C = 1/2R$, where R is the radius of curvature in the plane transverse to the beam. Curvature is signed according to the measured electric charge of the track. This is related to the transverse momentum, p_T , according to the formula, $p_T = B/2C$ where B is the magnitude of the magnetic field.
- ϕ_0 , the azimuthal angle at the point of closest approach in the transverse plane.
- Λ , equivalent to $\cot(\theta)$ where θ is the polar angle.

Silicon

The silicon detector is really a combination of three different subdetectors, Layer 00 (L00) [31], the Silicon Vertex Detector II (SVXII) [32], and the Intermediate Silicon Layers (ISL) [33]. The silicon detectors are essentially reverse-biased diodes. When

a charged particle strikes the strip, it ionizes the silicon creating an electron-hole pair. The voltage draws a current which is read out by chips mounted at the end the detectors. Although very precise, silicon detectors are not radiation hard and can be damaged easily.

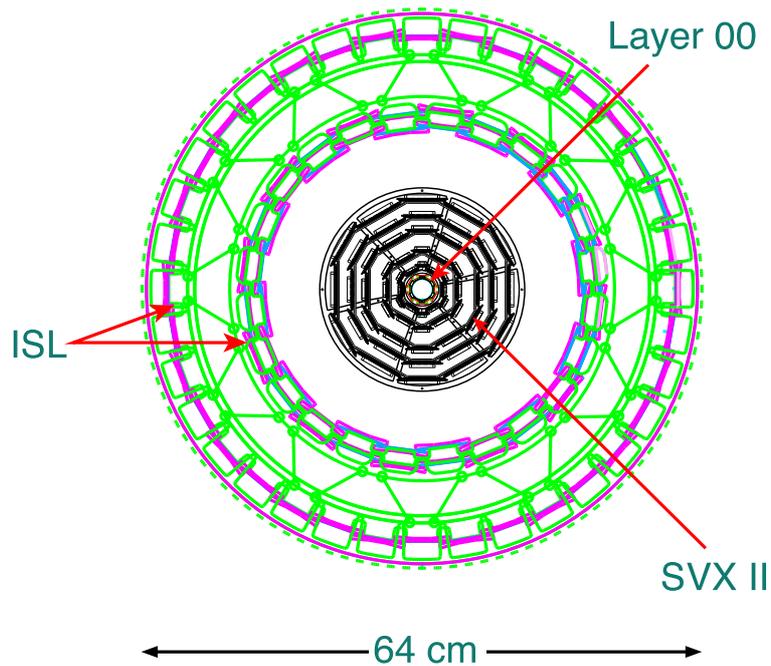


Figure 2.5: $r - \phi$ view of the L00, SVXII, and ISL subdetectors.

The SVXII is the primary subdetector of the silicon tracker. It is comprised of five concentric layers of double-sided silicon microstrip detectors (this is shown in Figure 2.5). The microstrips are arranged so that one side is aligned with the z -axis (axial) and the other side is at an angle (stereo). The stereo angles are either fully perpendicular (90°) or at small angles (1.2°). The SVXII is situated in three cylindrical barrels 32 cm in length and with radii extending from 2.5 cm to 10.6 cm. The barrels are subdivided into 12 azimuthal wedges of 30° each. The five layers

within each wedge are further subdivided into electrically independent modules called ladders, for a total of 360 ladders in the SVXII detector. In total, the SVXII has an area of 1.5 m^2 and nearly 270,000 channels. Including L00 and ISL, the entire silicon detector has $\sim 700,000$ channels.

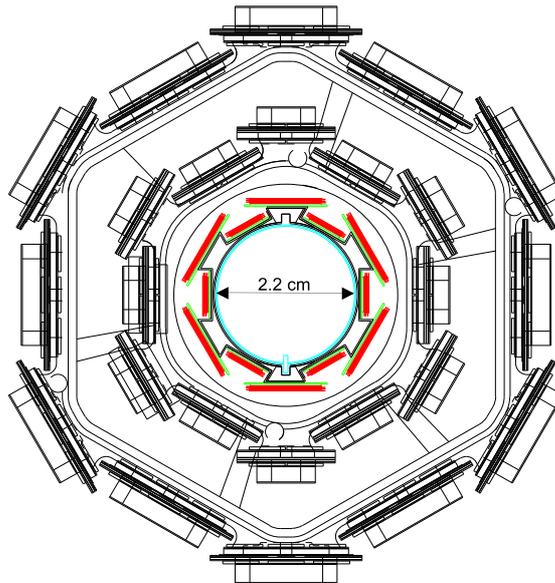


Figure 2.6: $r - \phi$ view of the L00 detector. Note the overlapping hexagonal structure.

L00 is a single-sided silicon microstrip detector mounted directly on the beampipe which has a radius of 1.2 cm. L00 is comprised of two overlapping hexagonal geometries with $r = 1.25 \text{ cm}$ and $r = 1.5 \text{ cm}$. This is shown in Figure 2.6. L00 improves the impact parameter resolution particularly of low p_T tracks.

The ISL is designed to extend the silicon coverage up to an η of 2.0. It is comprised of one layer (for $|\eta| < 1.0$) or two layers (for $1.0 < |\eta| < 2.0$) of double-sided silicon. The radii of the ISL extends from 20 cm to 28 cm.

Drift Chamber

The Central Outer Tracker (COT) [34] is a large, cylindrical open-cell drift chamber fully covering the $|\eta| < 1.0$ region and partially covering up to $|\eta| < 2.0$. The COT is divided into 8 superlayers consisting of 12 sense wires per layer, for a total of 96 layers. Within a superlayer, sense wires and field shape wires alternate. Superlayers are subdivided by cathode field panels, made of 0.25 mil Mylar and 450 Å gold plating. This configuration is shown in Figure 2.7. Superlayers alternate between axial and stereo (2°) measurements.

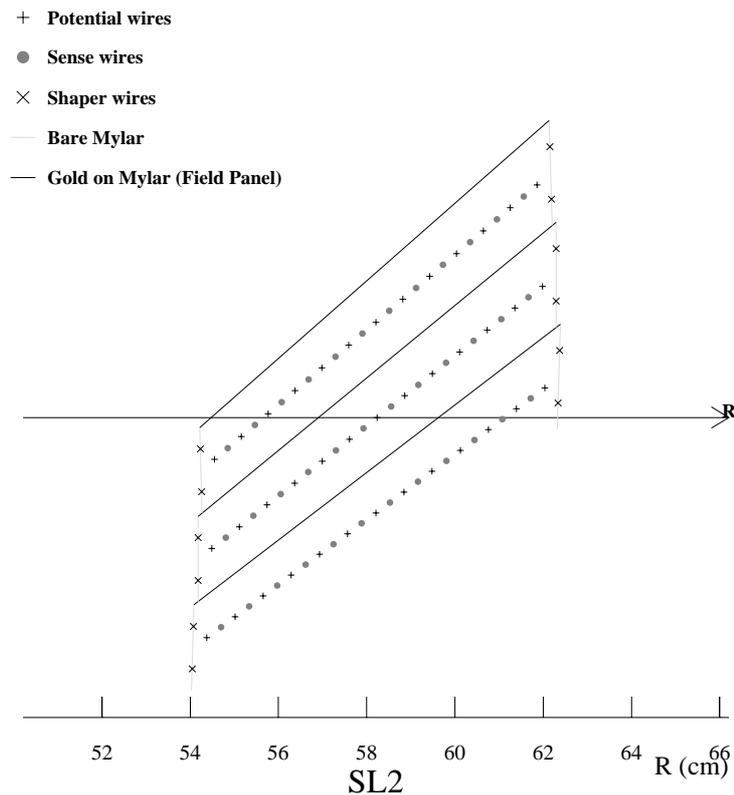


Figure 2.7: Illustration of a cell in superlayer 2 of the COT.

The COT is filled with a nearly equal mixture of Argone and Ethane gas, with a small $\sim 1\%$ component of isopropyl alcohol. Around the summer of 2004, a small component of oxygen was also added to prevent hydrocarbon chains from building up on the wires of the COT.

As charged particles pass through the COT, they ionize the gas. The resulting electrons drift toward the sense wires placed at high voltage. As the electrons get near the wire they avalanche, giving a charge pulse. The timing of these COT hits and the total charge deposited, encoded as a pulse width, are measured and corrected. The maximum drift time for this configuration is 177 ns, which prevents event pileup from different $p\bar{p}$ bunch collisions. Fast readout allows for the quick reconstruction by the trigger system.

2.2.2 Calorimetry

Calorimetry at CDF has a projective tower geometry and is divided in electromagnetic (EM) and hadronic (Had) components. In both cases, the calorimeters are sampling calorimeters, sandwiches of alternating lead (EM) or steel (Had) and plastic scintillator. Scintillator light is guided into wavelength shifters and then to Hamamatsu PMTs which are subsequently amplified and read out. The material induces photons to convert ($\gamma \rightarrow e^+e^-$), electrons to bremsstrahlung, and hadronic particles to undergo a nuclear interaction. These interactions cascade resulting in electromagnetic and hadronic showers in the respective calorimeters. An illustration of a central calorimeter tower is shown in Figure 2.8.

The central EM calorimeter (CEM) [35] is composed of towers segmented into

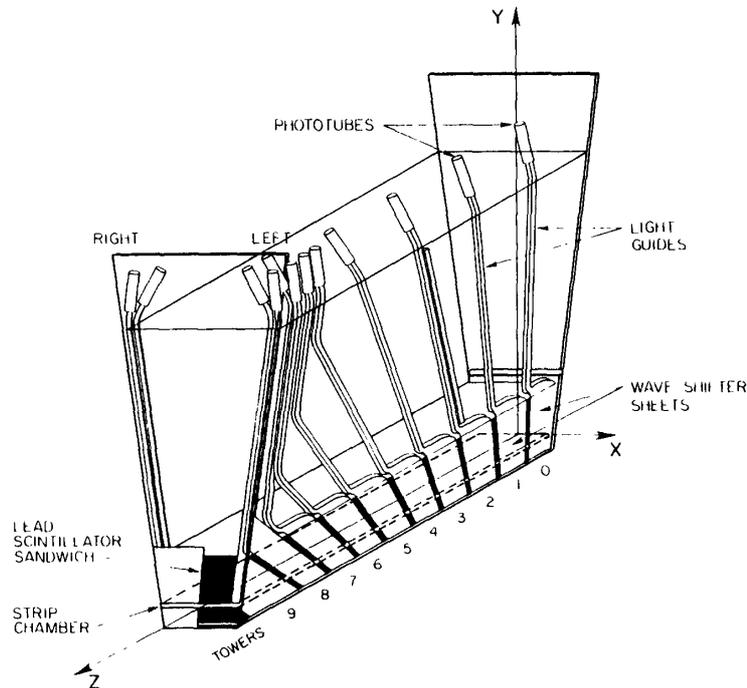


Figure 2.8: Illustration of a calorimeter tower.

$\Delta\eta = 0.11$ and $\Delta\phi = 15^\circ$, covering a range up to $|\eta| < 1.1$. The CEM is approximately 18 radiation lengths (X_0) and less than one interaction length (λ_I) deep. The resolution is $\sigma/E = 13.5\%/\sqrt{E_T} \oplus 2\%$ (E_T measured in GeV), dominated by the stochastic term at low energy and the noise term at high energy.

The central hadronic calorimeter (CHA) [36] has the same projective tower geometry, but only covers $|\eta| < 0.9$. The CHA is $4.7 \lambda_I$ deep, and has a resolution of $\sigma/E = 75\%/\sqrt{E_T} \oplus 3\%$. The wall hadronic calorimeter (WHA) is the same as the CHA but covers the gap between $0.7 < |\eta| < 1.2$ (see Figure 2.3 for an illustration of this geometry).

The plug region covers the $|\eta|$ range from 1.1 to 3.6. The plug electromagnetic

calorimeter (PEM) is $23.2 X_0$ and the plug hadronic calorimeter (PHA) is $6.8 \lambda_I$.

Embedded in the CEM is the central shower max detector (CES), wire and strip chambers located at approximately shower maximum $\sim 6X_0$. A similar instrument, the PES, is embedded in the PEM. The wire chambers run along the z axis and measure the shower profile in the $r - \phi$ view, and the strip chambers run along the $r - \phi$ axis and measure the shower profile in the z view. The purpose of the CES is to provide a high resolution reconstruction of the transverse profile of the electromagnetic shower. The wires and strips are finely segmented and provide a positional resolution $\sim 2 - 3$ cm.

2.2.3 Muon Chambers

The muon chambers sit outside of the calorimeters and are separated by a thick layer of steel shielding. The muon chambers are layers of single wire drift chambers with a gas composition similar to the COT. Scintillators paired with the drift chambers are used for timing and cosmic ray veto information. Muons minimum ionize through the calorimeter, neither showering like electrons nor having nuclear interactions like hadrons. Those with a sufficient range ($p_T \gtrsim 2.0$ GeV/ c) pass through the muon chambers where they are detected. A series of hits in the different layers of the muon chambers can be reconstructed into a muon ‘stub’ which can then be matched to a track (see Figure 2.9 for an illustration).

The primary subcomponents of the muon system relevant to this analysis are the central muon detector (CMU), the central muon upgrade (CMP), and the central muon extension (CMX). The CMU [37] has a cylindrical geometry and is built directly

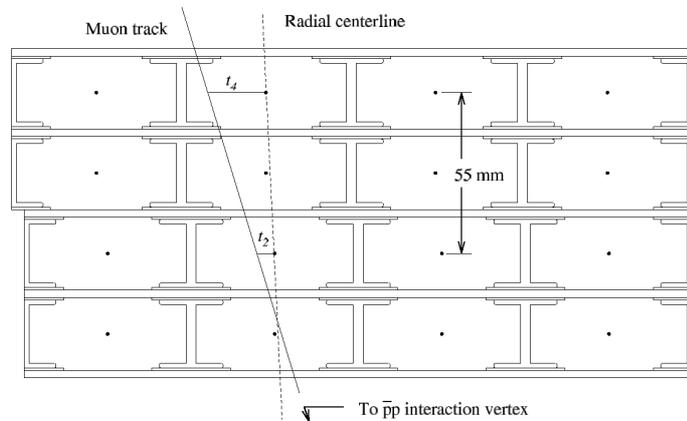


Figure 2.9: Illustration of a cross sectional view of the CMU, showing the path of a charged particle.

on the CHA, covering a pseudorapidity region up to 0.6. Because $\lambda_I \approx 5.5$ at this point, a significant amount of ‘punch-through’ - that is, hadrons which have managed to escape the hadronic calorimeter without fully depositing their energy - is present. Beyond the CMU is another $3.5 \lambda_I$ of steel (comprised from the return yoke for the solenoid) followed by the CMP. The CMP has a box geometry and covers the same η region as the CMU. By combining the CMU and the CMP, fake muons due to punch-through can be efficiently rejected. Both the CMU and the CMP consist of four layers of drift cells staggered by one-half cell per layer.

The CMX extends the muon η coverage from 0.6 to 1.0 (Figure 2.10 shows the coverage of the muon chambers). Eight layers of staggered drift tubes comprise the CMX. Its geometry is conical but gaps prevent a full 360° coverage in ϕ . The keystone and miniskirt are the segments of the CMX which sit on top and bottom, respectively, of the body of detector and help complete some of the ϕ coverage. The ϕ range for the keystone and miniskirt are $75 < \phi < 105$ and $225 < \phi < 305$, respectively, where

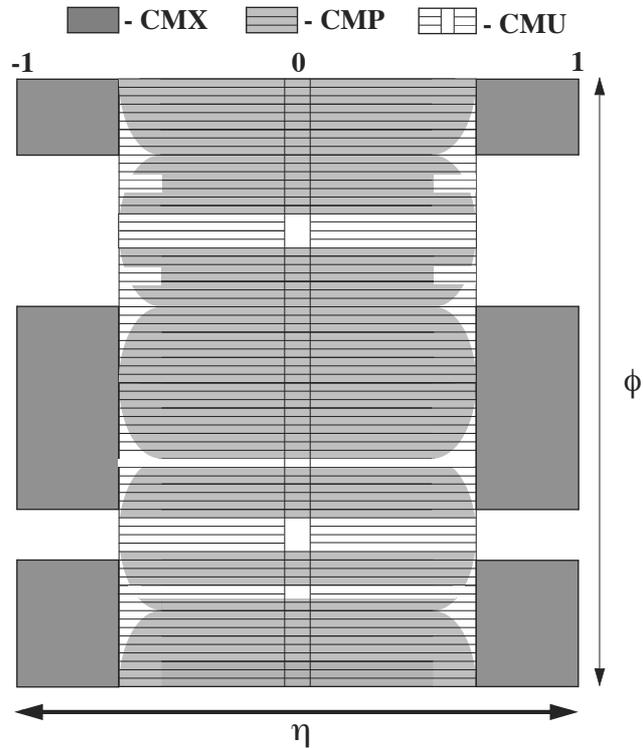


Figure 2.10: Central muon chamber coverage at CDF.

ϕ is measured in degrees.

2.2.4 Luminosity Monitors

The Čerenkov Luminosity Counters (CLC) is a collection of gaseous Čerenkov detectors designed to infer the instantaneous luminosity by measuring the number of particles resulting from inelastic $p\bar{p}$ scattering in the very forward region ($3.7 < |\eta| < 4.7$). The measured luminosity has a relative uncertainty of 5.9% dominated by the total inelastic $p\bar{p}$ cross section and the uncertainty on the average number of $p\bar{p}$ interactions per bunch crossing [38].

2.2.5 Trigger and Data Acquisition Systems

The data acquisition and trigger systems select events of “interest” to be recorded to tape. In Run 2, the collision rate is approximately the same as the crossing rate, ~ 2.5 MHz, which must be reduced to ~ 100 Hz for recording. CDF uses a three tiered trigger system (Levels 1, 2, and 3) with a sufficient rate reduction to allow for processing by the subsequent level with minimal deadtime. The Level 1, 2, and 3 triggers reduces the rate to approximately 20 kHz, 500 Hz, and 100 Hz, respectively. A block diagram of this procedure is shown in Figure 2.11. In order to accommodate the high trigger rates of certain processes, we ‘prescale’ the triggers taking only 1 out of every N events. Prescaling is done dynamically to address the luminosity dependence of the trigger rates.

The Level 1 trigger consists of custom designed hardware to make decisions based on simple physics quantities, such as electrons, photons, muons, the total transverse energy (E_T), and missing E_T . The eXtremely Fast Tracker (XFT) reconstructs tracks in the axial COT superlayers and provides geometric information for matching. Coarse track matching can be performed at Level 1 so that physics objects such as muons (which involve a track pointing to a muon stub) can be triggered on. Clustering of towers for jet identification is not performed until Level 2.

The Level 2 trigger uses dedicated hardware to do limited event reconstruction. Coarse information available at Level 1 is refined at Level 2. CES information is available at Level 2 as well, which aids electron/photon ID and reduces the combinatoric background for EM shower/track matching. Another important contribution to Level 2 is the secondary vertex tracker (SVT) which is the first trigger at a hadron

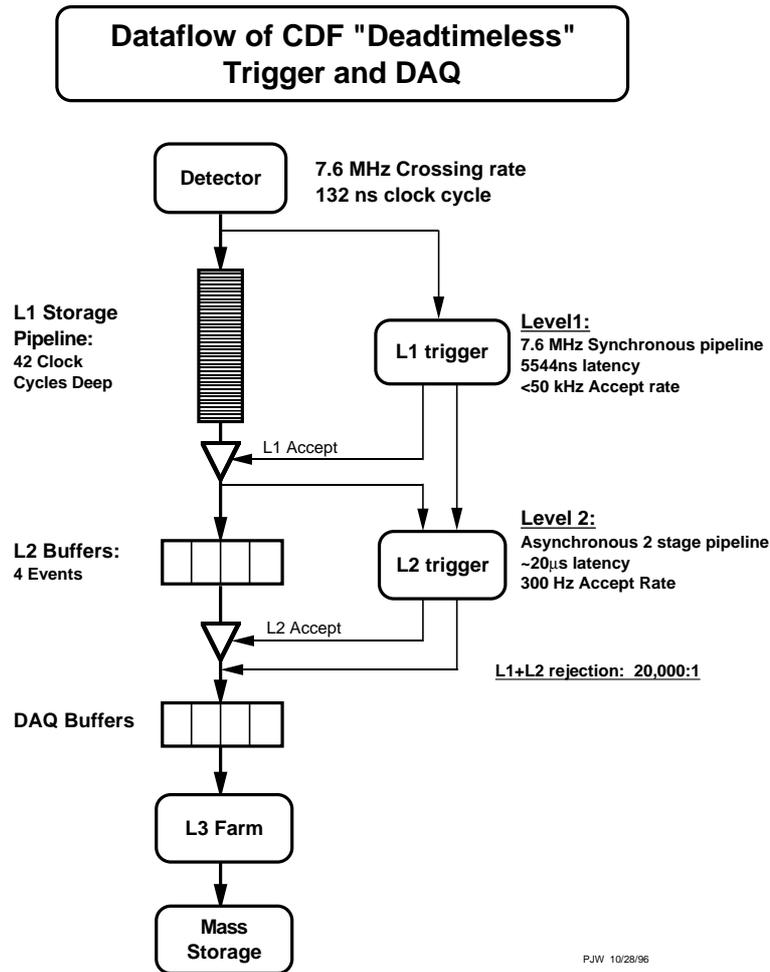


Figure 2.11: A block diagram of the CDF data acquisition system.

collider to trigger on tracks with a high impact parameter. The availability of this trigger aided in the first observation of B_s oscillations at CDF.

The Level 3 trigger has two components, an 'event builder' that uses custom hardware to assemble data from all the CDF subdetectors and a computer farm to fully reconstruct events. The Level 3 software is similar to the offline reconstruction software and can make sophisticated decisions about particle ID or other physics

objects. Those events which pass the Level 3 requirement are separated into event streams and are committed to permanent storage.

The High p_T lepton triggers used in this analysis fire on central electrons in the CEM and central muons in the CMUP or CMX with $p_T > 20$ GeV/ c . These triggers are not prescaled. The central electron trigger fires at an average rate of approximately 1 Hz, and the central muon triggers fire at a combined average rate of approximately 0.5-0.7 Hz.

Chapter 3

Event Reconstruction

In the previous chapter, we discuss the experimental apparatus needed to produce and measure $t\bar{t}$ events. In particular, a three tiered trigger system is required to select events of interest. The trigger system uses a combination of hardware and software techniques to reconstruct physics objects *online*. In this chapter we discuss the *offline* reconstruction of physics objects by sophisticated software algorithms once they have been recorded to tape. We discuss tracking, lepton identification, calorimeter objects, and vertex reconstruction.

3.1 Tracks

Tracks are reconstructed using hits in the COT and silicon detectors. Two algorithms are run in parallel to reconstruct tracks: Segment Linking (SL) and Histogram Linking (HL). The SL algorithm forms track segments in each of the superlayers individually and tries to link them together, starting from the outermost superlayers.

By contrast, the HL algorithm begins with a segment and the beam position which defines a nominal curvature radius. The distance between hits and the nominal road are binned in a histogram. The most populated bin indicates a potential track. Duplicate tracks from the SL and HL algorithms are removed, and preference is given to the SL algorithm. Track candidates are fit to a helix, and basic quality requirements are applied. If a hit associated with a track has a residual > 600 microns, then the hit is dropped and the track refit again.

At CDF, the final tracks used in physics analyses are determined by a wide range of algorithms. Typically tracks are formed by taking candidate COT tracks and extrapolating them into the silicon detector. Silicon hits that are within a road are added to the track. To recover forward tracks that do not traverse much of the silicon, an algorithm which begins inside and extrapolates out is also performed. COT tracks with no silicon are classified as COT-only.

In general, if the silicon detector is used, these tracks must have ≥ 3 axial silicon hits. If the COT is used, ≥ 2 axial and ≥ 2 stereo superlayers must be used with ≥ 5 hits in each superlayer. Further criteria are typically imposed depending on the type of algorithm used.

The tracking algorithm is 98-99% efficient for tracks with $p_T > 1$ GeV/ c . COT-only tracks have a transverse momentum resolution, $\sigma_{p_T}/p_T \approx 0.15\% \times p_T$. When combined with hits from the SVX and ISL, the transverse momentum resolution is $\sigma_{p_T}/p_T \approx 0.07\% \times p_T$.

3.2 Leptons

In this section we discuss the identification of electrons and muons which form the basis of the *lepton+jets* dataset. The leptons are required to be consistent with those produced from W decay: high p_T and isolated. We discuss the identification of soft electrons embedded in jets in later chapters.

3.2.1 Electrons

Central, high E_T electrons are identified by matching tracks to electromagnetic calorimeter clusters. Electrons are selected with the following criteria:

- Fiducial to the CES and CEM: The CES and CEM (in conjunction with the trackers) are the primary means by which central electrons are distinguished from other physics objects.
- $E_T > 20$ GeV: E_T is calculated by measuring the total EM energy deposition in a electromagnetic cluster (corrected for non-linearities and tower-to-tower variations) and multiplying by the sine of the polar angle of the associated track: $E_T = E \sin(\theta)$. The cluster used for this variable is a two-tower cluster, where the towers are chosen to be the closest two η -adjacent towers to the track. Since electrons bremsstrahlung easily, the energy, rather than the momentum, is typically used to characterize electrons.
- $p_T > 10$ GeV: If silicon hits are not present, the track parameters are calculated assuming that it originated from the beamline (see Section 3.4.1 for beamline reconstruction).

- ≥ 3 Axial COT SLs with ≥ 5 hits per SL and ≥ 2 Stereo COT SLs with ≥ 5 hits per SL: This ensures that the COT track is well measured.
- Track $|z_0| < 60$ cm.
- Not a conversion: Electrons that result from conversion of a photon into e^+e^- pairs cannot be differentiated from W or Z electrons on the basis of more stringent electron selection requirements. Instead, a geometric technique is used to identify conversion electrons. The event is searched for an oppositely-charged track with $|\Delta \cot(\theta)| < 0.04$ and $|sep| < 0.2$ cm. sep is defined as the perpendicular distance between two tracks when they are parallel. If a third track forms a conversion pair with either the primary electron candidate or the conversion partner, then the event is no longer vetoed. This requirement is implemented to allow for “trident” electrons, where an electron bremsstrahlung a photon which promptly converts.
- Calorimeter isolation < 0.1 : The isolation used here is defined as the ratio of the total hadronic and electromagnetic energy in a cone of $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 0.4$ not in the electron cluster to the total energy in the cluster. This is a calorimeter-based isolation measurement, different from the track-based isolation measurement we will use later for the identification and measurement of soft electrons.
- $E_{Had}/E_{EM} \leq 0.055 + 0.00045 \times E$ (GeV): This requires that the shower is mostly electromagnetic and suppresses hadronic objects. The second term accounts for some leakage in the CEM when the electron is especially energetic. For this

variable, a three-tower calorimeter cluster is used.

- $L_{shr} \leq 0.2$: L_{shr} characterizes the lateral sharing of the electromagnetic shower between calorimeter towers. Although the Molière radius¹ is small (~ 2 cm) for the CEM, the total content of an electromagnetic shower can be spread between different CEM towers adjacent in η , particularly when the electron track is not perfectly aligned the projective geometry. The value of L_{shr} is the sum over towers of the difference between the expected and measured energy deposits divided by the root-mean-squared uncertainty. L_{shr} uses a three tower cluster.
- $E_{EM}/p \leq 2$ unless $p_T \geq 50$ GeV/ c : The total electromagnetic energy of the cluster must be consistent with the momentum of the track. Both detector resolution and bremsstrahlung can affect the ratio. Very high momentum electrons have this requirement released.
- $|\Delta_{strip}| \leq 3$ cm and $-3.0 \leq Q\Delta_{wire} \leq 1.5$ cm: The distance between the track extrapolation and the center of energy of the electromagnetic cluster in the CES must be small. The distance is calculated only in the plane of the respective detector (strip or wire).
- $\chi^2_{strip} \leq 10$: The transverse profile of the electromagnetic cluster in the CES strip chamber must be consistent with an electron. The measured shape is compared to that of a test beam of 10 GeV electrons with a χ^2 test. The χ^2 value is also scaled to account for variations of the profile as a function of energy

¹The Molière radius is a characteristic constant of material describing the transverse radius of an electromagnetic shower, approximately equal to $0.027X_0(Z + 1.2)$, where X_0 is the radiation length and Z is the atomic number of the material.

and momentum.

The efficiency of these cuts is approximately 79%, using a sample of electrons from Z decay. The ratio of the efficiency in data and Monte Carlo - *i.e.* the data-MC efficiency scale factor (SF) - is approximately 0.98. These values vary slightly with time due to varying instantaneous luminosity levels and detector effects.

3.2.2 Muons

Central high p_T muons are identified by matching tracks to muon stubs. Additionally, the associated calorimeter deposition must be consistent with a minimum ionizing particle. Since the CMU has a large punch-through rate (see Section 2.2.3), the CMP is added to suppress fake muons. Two classes of muons are considered: those with stubs in both the CMU and CMP, called CMUP muons, and those with a stub in the CMX detector. Muons are selected with the following criteria:

- Fiducial to the CMU/CMP or CMX.
- $p_T > 20$ GeV/ c .
- $E_{EM} < 2$ GeV for $p < 100$ GeV/ c and $E_{EM} < 2 + 0.0115 \times (p - 100)$ GeV: The electromagnetic content of the towers associated with the muon track must be small. The requirement is loosened when the muon is very high p_T to accommodate the growing energy deposition. p is the momentum measured in units of GeV/ c .
- $E_{Had} < 6$ GeV for $p < 100$ GeV/ c and $E_{Had} < 6 + 0.028 \times (p - 100)$ GeV: Similarly, the hadronic content must also be small. Again, the p above is the

momentum measured in units of GeV/c .

- Calorimeter Isolation < 0.1 : The isolation defined here is the total E_T in a cone of $\Delta R \leq 0.4$ divided by the transverse momentum.
- ≥ 3 Axial COT SLs with ≥ 5 hits per SL and ≥ 2 Stereo COT SLs with ≥ 5 hits per SL: This ensures that the COT track is well measured. This same requirement is used for electrons.
- Track $|z_0| < 60$ cm.
- $|d_0| < 0.02$ (0.2) cm when the tracks use (do not use) silicon information. The impact parameter is corrected to be relative to the beamline. This requirement suppresses cosmic ray muons.
- $|\Delta_x| < 7$ cm (CMU), 5 cm (CMP), 6 cm (CMX): the distance between the extrapolated track position and muon stub in the chambers must be small. Multiple scattering and the intrinsic resolution of the chambers account for the difference.
- $\rho_{\text{COT}} > 140$ cm (CMX muons): the exit radius of the track from the COT for CMX muons must be greater than 140 cm. This ensures that the track is well measured and fiducial.
- Track fit $\chi^2 < 2.3$ (2.75): The χ^2 of the track fit must be good in order to suppress decays-in-flight (*e.g.* $K \rightarrow \mu\nu$, $\pi \rightarrow \mu\nu$) which contribute very high p_T fake muons. Two different values for the requirement are used in different

time periods, but both are $\sim 99\%$ efficient. The χ^2 must be less than 2.3 for the first 300 pb^{-1} of data, and 2.75 for the rest.

The muon identification efficiency, measured in a sample of $Z \rightarrow \mu^+\mu^-$, is approximately 90% and 92% for CMUP and CMX muons, respectively. The data-MC muon identification efficiency scale factor is approximately 0.98 and 1.00 for CMUP and CMX muons, respectively. These values vary slightly ($\sim 1\%$) with time. The CMX efficiency also changes slightly if the arches, miniskirt, or keystone is used. The precise values and uncertainties of the efficiencies and SFs for electrons and muons will be revisited in Chapter 9.3.3.

3.3 Calorimeter Objects

In this section, we discuss the reconstruction of calorimeter objects, specifically jets and missing energy.

3.3.1 Jets

Jets are broad streams of particles resulting from quark or gluon hadronization. For this analysis, we use the `JetClu` algorithm to identify jets by a fixed-cone clustering technique. The algorithm is seeded by a local maximum in the total tower energy (with at least 1 GeV of deposited energy) and considers the energy deposits in all towers in a fixed $\Delta R \leq 0.4$ cone. The algorithm recalculates the E_T -weighted center of the cluster, and reiterates over the procedure until the jet energy and its center is stable. Jets which share more than 50% of their energy are merged together.

The energy of the jet must be corrected for a variety of effects. The initial level of corrections accounts for tower-to-tower variations and non-linearities in the PMT response to charge depositions. The second level of corrections adjusts for the η dependence of the detector response, which results from gaps in coverage and varying responses from the different calorimeters. This correction is determined by reconstructing ‘di-jet’ events. On average, the total energy of each jet should be balanced by the other, allowing the total calorimeter response to be normalized to the $0.2 < |\eta| < 0.6$ region.

The third level corrects for multiple interactions per beam crossing which could deposit energy in the calorimeters. The overall contribution is subtracted off on average, by examining minimum-bias events and parameterizing the correction as a function of the number of reconstructed vertices.

The last level of corrections used for this analysis adjusts the jet energies to account for gaps in instrumentation and non-linearities in the total calorimeter response. This creates an absolute scale of comparison between different experiments, for instance. This correction is extracted from Monte Carlo.

Other corrections attempt to estimate the absolute energy of the parton responsible for the jet. These corrections address energy that escapes the cone and extra particles from the underlying event which could find their way into the calorimeter.

The fractional systematic uncertainty on the jet energy corrections for each component is shown in Figure 3.1. In this analysis, we do not use the ‘out-of-cone’ or ‘underlying event’ corrections.

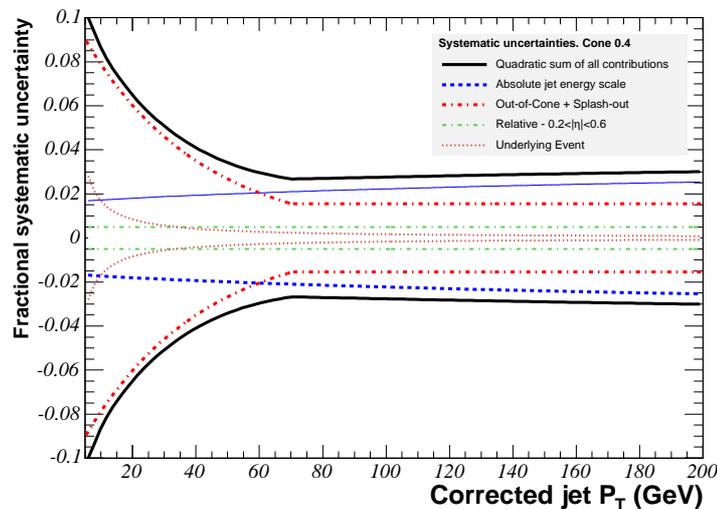


Figure 3.1: Fractional systematic uncertainty as a function of jet p_T . We do not use the ‘out-of-cone’ or ‘underlying event’ corrections.

3.3.2 Missing Energy

Missing energy is a signature for neutrinos and other exotic particles that do not interact with the detector. It is calculated by integrating over all calorimeter towers dotted with their radial direction from the beam axis to the tower:

$$\vec{E}_T = \sum E_i \sin \theta_i \hat{n}_i. \quad (3.1)$$

This definition includes un-clustered energy in the event. The missing transverse energy, \cancel{E}_T , is $-\vec{E}_T$.

If a high p_T muon is reconstructed in the event, the \cancel{E}_T is corrected since muons are typically minimum ionizing and do not deposit their energy into the calorimeter. The \cancel{E}_T is also corrected for the jet energy corrections mentioned above and the location of the primary vertex, discussed below.

3.4 b -Tagging

Identifying bottom and charm jets - typically referred to as heavy flavor (HF) - is an important task for $t\bar{t}$ event reconstruction. b -tagging enhances the $t\bar{t}$ content relative to the backgrounds, increasing the signal purity and significance. In this section, we briefly discuss the identification of b -jets using secondary vertex reconstruction. In the following chapters, we discuss b -jet identification using soft electron tagging.

3.4.1 Beamline and Primary Vertex Reconstruction

The beamline is reconstructed offline on a run-by-run² basis using two different algorithms. Each algorithm runs in both silicon+COT and COT-only modes. The first algorithm measures the average d_0 and ϕ_0 calculated with respect to $(0,0)$ of tracks. If the beamline is offset from $(0,0)$, its position will evince itself as a sinusoidal dependence between d_0 and ϕ_0 . Fitting the shape determines the x and y position of the beamline. Measuring the x and y position in different segments along z reconstructs the full beamline.

The second algorithm reconstructs the primary vertex position. From the distributions of the reconstructed primary vertices, one can directly derive the beam position and other beam parameters. The Primary vertex is reconstructed using the average z position of tracks as a seed along the beam direction. A constrained fit determines the most likely position of the primary vertex. An iterative procedure of adding or removing tracks with a good or bad χ^2 fit and recalculating the primary

²If a run is particularly long, the run is divided into a run-section, and the beamline is calculated separately by section.

vertex position is done until the position is stable. If multiple primary vertices are reconstructed, then the one with the set of tracks with the largest total p_T is used.

3.4.2 Secondary Vertex Reconstruction

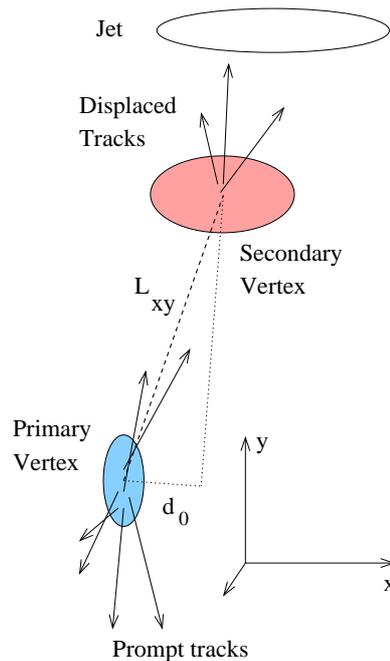


Figure 3.2: An illustration of a primary and secondary vertex.

The primary b -tagging algorithm used at CDF is `SECVTX`, which looks for two or more tracks in a jet forming a secondary vertex. An illustration of this is shown in Figure 3.2. The algorithm begins with well-measured tracks with silicon hits and a large impact parameter significance with respect to the reconstructed primary vertex. For a first pass, vertices are reconstructed out of pairs of tracks. Those vertices with a good fit χ^2 are kept, and `SECVTX` attempts to add more tracks to the vertex. If no secondary vertex is found, a second pass is made where a vertex fit is attempted using

a more restricted subset of the tracks but a less restrictive vertexing requirement.

Two different operating points of the SECVTX algorithm are used, commonly referred to as “tight” SECVTX and “loose” SECVTX. Tight SECVTX has a lower tagging efficiency but also a lower mistag rate, that is the rate at which light jets are mis-identified as heavy flavor. The efficiency and mistag rate as a function of Jet E_T are shown in Figure 3.3. The ratio of the tagging efficiency in data to MC is measured to be 0.95 ± 0.05 .

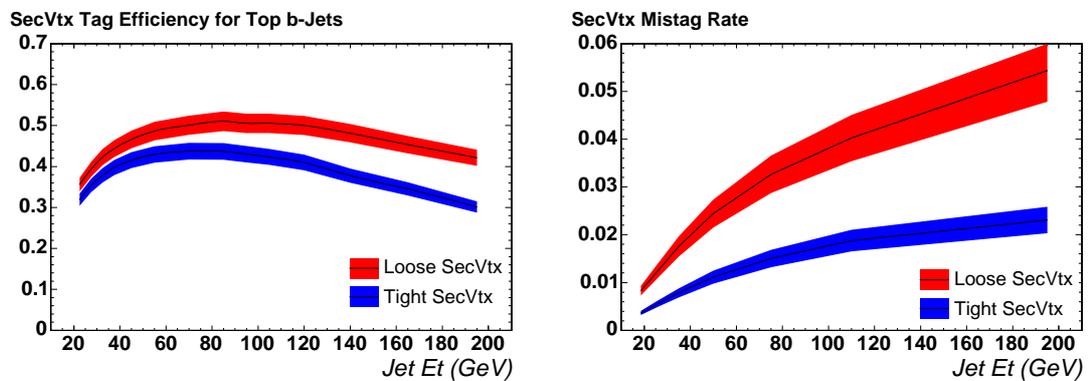


Figure 3.3: The tagging efficiency and mistag rate for loose and tight SECVTX as a function of the jet E_T .

Chapter 4

Soft Electron Tagging

4.1 Introduction

In this chapter, we present a soft electron tagger (SLT_e) to identify electrons from the semileptonic decay of bottom and charm hadrons as well as other sources. In this introduction, we present an outline of our strategy to develop and implement the SLT_e .

The identification of heavy flavor (HF) jets via the tagging of a soft lepton embedded within the jet played a role in the top quark discovery at CDF and DØ in Run 1 [10, 11]. This tagging method is possible because the semileptonic branching fraction of a bottom or charm hadron is approximately 10% per lepton (either electron or muon). For bottom hadrons, an additional 10% can be recovered from the subsequent semileptonic decay of a daughter charm hadron. From these branching fractions, we expect approximately 68% of $t\bar{t}$ events in the *lepton+jets* channel to contain a soft electron or muon. This technique complements other taggers such as

SECVTX that instead rely on the long lifetime of heavy flavor jets.

The SLT_e algorithm uses the tracker, central calorimeter and, in particular, the shower maximum chambers (CES) to identify electrons embedded in jets. The tagging algorithm is “track-based” (as opposed to “jet-based”) in that we consider every track in the event a tag candidate. In order to understand the behavior of electrons within jets in the detector, we use electrons that originate from photon conversions as a control sample. Two reasons motivate this choice. First, it is simple to construct a large sample of conversion electrons with high purity and which does not suffer from biases due to the trigger selection. This is done by triggering on a 8 GeV electron and looking for a conversion partner. Second, the p_T spectrum of HF electrons in $t\bar{t}$ events is very broad, but - like conversion electrons - is dominant at low p_T . This feature is shown in Figure 4.1 which compares the p_T of electrons from bottom decay, charm decay, and photon conversions in $t\bar{t}$ MC events. The understanding of the low p_T behavior of electrons therefore is critical to HF electron tagging. The high p_T behavior of electrons is extrapolated from low p_T and validated with a high p_T sample of electrons from Z bosons.

The two primary backgrounds to HF electrons are fake electrons (typically charged pions, kaons, and protons) and real electrons from conversions that are embedded in jets. Electrons from Dalitz decay of π^0 is a small, but non-negligible, source of background as well. Tracks that fake electrons are modeled with generic tracks from events triggered on a high E_T jet. The SLT_e algorithm rejects those tracks by identifying electromagnetic showers in the calorimeter that have a transverse shape and energy distribution consistent with a single electron in a hadronic environment.

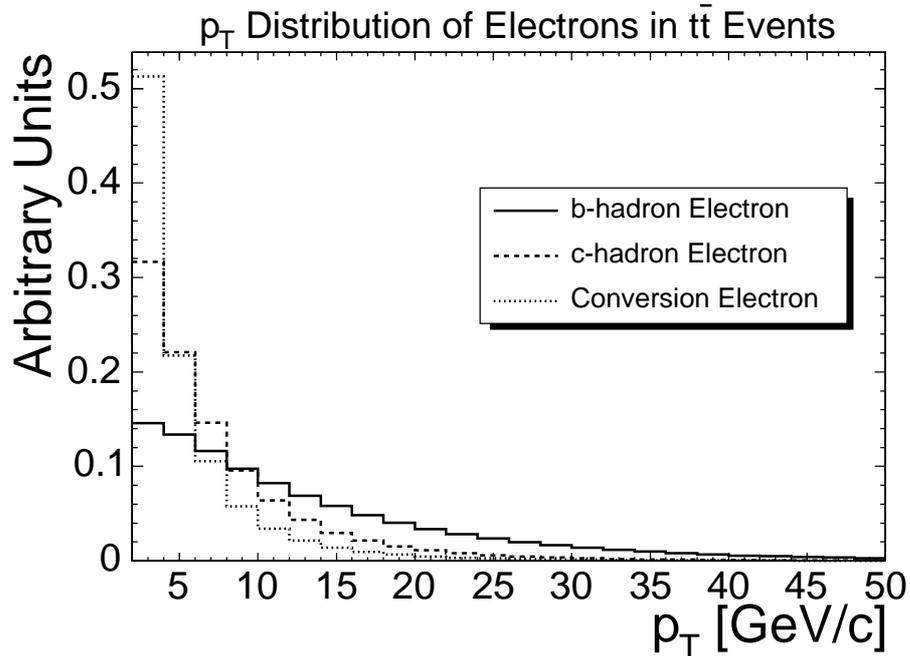


Figure 4.1: p_T spectrum of electrons from b hadron decay, c hadron decay, and photon conversion in $t\bar{t}$ MC events (generated with $M_{top} = 175 \text{ GeV}/c^2$). Distributions are normalized to unit area to show the shape.

The use of generic tracks in jets to model fake electrons is motivated by their kinematic and environmental similarity with fake electrons in $t\bar{t}$ and other high p_T samples. The drawback of this sample is that, whereas only $\sim 1\%$ of the generic tracks are electrons, these electrons constitute a significant fraction of the SLT_e tags.

It is important to keep in mind that while we have employed conversion electrons as our signal model, they actually constitute one of the primary backgrounds to HF electron tagging. Conversion electrons are identified with tracking information by looking for a conversion partner track somewhere in the event. If no partner is found, a conversion electron can also be rejected if the candidate track is missing too many silicon hits. Tracks from b -jets are largely unaffected by this filter since - even in

energetic $t\bar{t}$ events - most b -jets decay within 2 cm of the primary vertex, and nearly all of them within 3 cm. The first layer used to identify conversion electrons, Layer 0 of the SVX, is located 2.5 cm from the center of the detector, (0, 0).

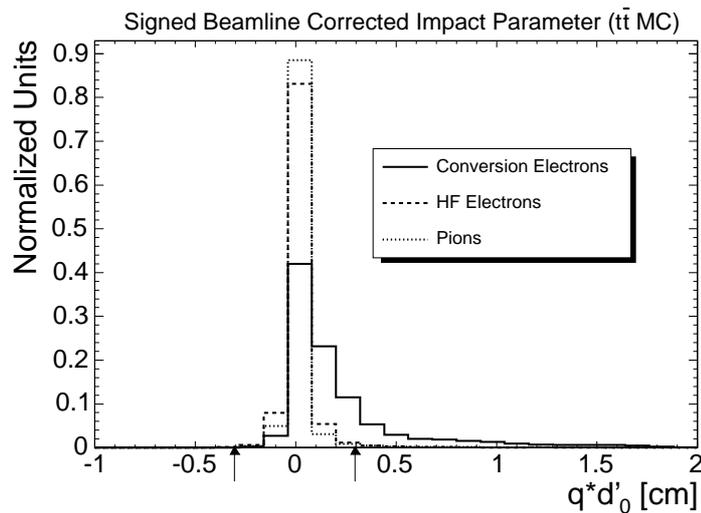


Figure 4.2: $q d'_0$ for electrons from conversions, electrons from heavy flavor, and pions in $t\bar{t}$ MC events (generated with $M_{top} = 175$ GeV/ c^2). Distributions are normalized to unit area to show the shape.

The tagger algorithm begins by selecting candidate tracks with p_T above 2 GeV/ c and $|d'_0| < 0.3$ cm, where d'_0 is the beamline corrected impact parameter. The impact parameter requirement is designed to reduce the conversion electron background as well as mis-measured tracks. Shown in Figure 4.2 is the d'_0 distribution signed by the charge of the track, q . For conversion electrons, this distribution is highly asymmetric about $d'_0 = 0$.

Once selected, the track is matched to a high E_T jet, and conversion candidates are rejected. The track extrapolated is used as a seed for electromagnetic clusters in the calorimeter and CES. We select calorimeter depositions based on distributions

from MC that are consistent with the electromagnetic shower of an electron. Candidate tracks which meet these requirements have their variables put into a likelihood discriminant that is based on templates built from conversions and generic tracks. A candidate SLT_e track that has a likelihood consistent with an electron is considered *tagged*. An illustration of HF electrons, conversion electrons, and fake electrons embedded in jets is shown in Figure 4.3.

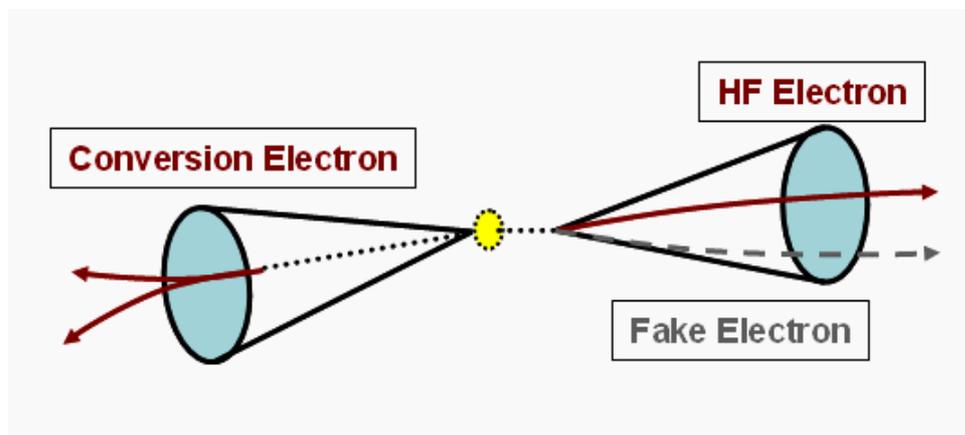


Figure 4.3: An illustration of HF electrons, conversion electrons, and fake electrons embedded in jets.

Because the CES variables are not well-modeled in MC, we parameterize the tagging efficiency of electrons and non-electrons in data. Applied to MC, this tagging efficiency functions as the probability of the SLT_e to tag a particular track. The tagging efficiency parametrization of real electrons is called the *tag matrix* and is measured with the conversion electron sample. The tagging efficiency parametrization of fake electrons is called the *fake matrix* and is measured with the generic tracks sample. The tag matrix must be corrected because the conversion electrons in this sample are typically more isolated than HF electrons. The fake matrix must be

corrected because of the real electron content of the generic jet sample. The efficiency of the conversion filter is also different in MC than in data. Data-MC scale factors (SFs) are measured to address the discrepancy.

In this chapter, we present the SLT_e tagging algorithm. In Chapter 5 we present the conversion electron filter and measure its efficiency and over-efficiency in various datasets. The data-MC efficiency and over-efficiency SF is also calculated there. In Chapters 6 and 7 we present the tag matrix and fake matrix, respectively, which predict the tagging rate for electrons and non-electrons in MC. In Chapter 8 we present a complete cross check of all of these elements in a high purity $b\bar{b}$ sample.

4.2 Sample Selection

Here we describe the collection of two control samples to understand the behavior of real electrons and fake electrons in the calorimeter and CES. These samples are also used as templates for the construction of a likelihood described in Section 4.3.5.

4.2.1 Pure Conversion Selection

The behavior of HF electrons is modeled with a pure sample of conversion electrons. This sample is collected through an inclusive 8 GeV electron trigger. We identify trigger electrons offline with selection criteria consistent with the Level 3 trigger criteria. Table 4.1 lists the criteria used (see Section 3.2.1 for the definition of the variables), and Figure 4.4 shows the corresponding distributions. Once the trigger leg is identified, we find the conversion partner by looking for a second, oppositely-charged track in the event close to the trigger electron. Since the partner leg rarely

($\sim 1\%$) fires the trigger, its electron identification quantities are unbiased. We use this partner leg to model HF electrons.

Trigger Leg Selection Criteria
Fiducial to CES
$ q\Delta_{wire} < 3.0$ cm
$ \Delta_{strip} < 5.0$ cm
$0.5 < E_{EM}/P < 2.0$
$E_{Had}/E_{EM} < 0.05$
$L_{shr} < 0.2$
$\chi_{strip}^2 < 10$
$E_T > 9.0$ GeV
$P_T > 8.0$ GeV/c

Table 4.1: Selection requirements for the 8 GeV trigger electron and primary leg of the conversion. See Section 3.2.1 for the definition of these variables.

We require that the partner leg be a well-measured track - that is, it must have ≥ 2 axial ≥ 2 stereo COT segments with ≥ 5 hits per segment - and fiducial to the CES. Fiduciality for electrons requires that the track must extrapolate to within 21 cm of the center of the CES in x and between 9 cm and 230 cm in z (see Figure 2.8 for the local coordinate system). The track must also not extrapolate to the chimney towers or the last tower of any central calorimeter module. For this sample, we also require that the partner leg does not extrapolate to the same towers as the trigger leg to prevent the partner's electron id variables from being affected by the other shower. The p_T threshold is set to 2 GeV/c, and the track must have a beamline corrected impact parameter, d'_0 , less than 0.3 cm. The track must also have a lower p_T than the trigger leg. This condition enforces that only one electron per conversion enters our sample, and it uniquely specifies the trigger leg. Finally, we use the track extrapolation position as a seed for wire and strip clusters in the CES. The tracks

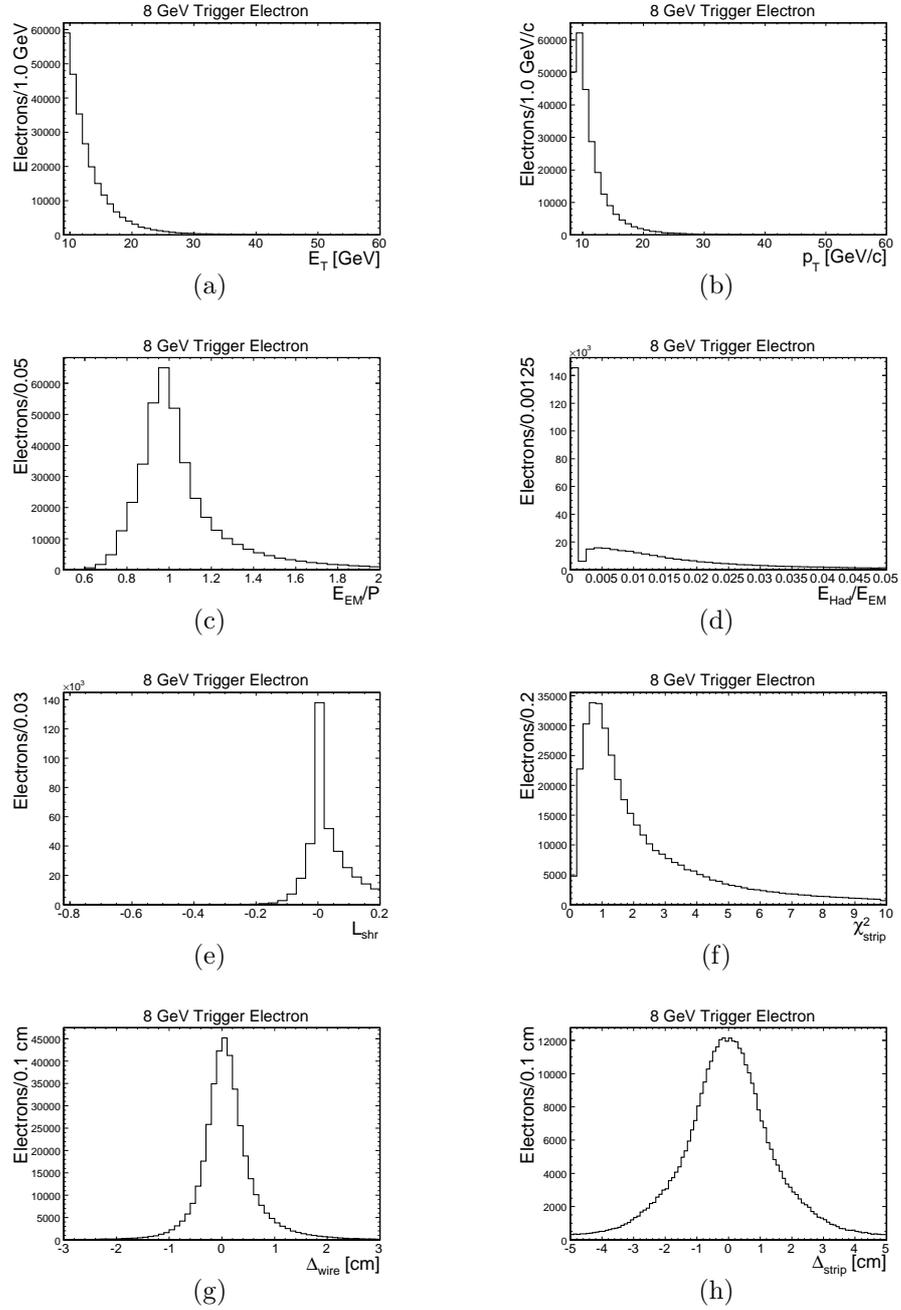


Figure 4.4: Distributions of the selection requirements for the 8 GeV trigger electron. Shown are E_T (a), p_T (b), E_{EM}/p (c), E_{Had}/E_{EM} (d), L_{shr} (e), χ^2_{strip} (f), Δ_{wire} (g), and Δ_{strip} (h). The distributions are calculated before any requirement on a partner leg is made.

Partner Leg Selection Criteria
Fiducial to CES
Seeds track-based CES cluster
COT Axial segments ≥ 2 (≥ 5 hits per segment)
COT Stereo segments ≥ 2 (≥ 5 hits per segment)
$ z_0 < 60$ cm
$p_T > 2$ GeV/c
$ d'_0 < 3$ mm
$ \eta < 1.2$
points to different tower than trigger leg
$p_T < p_T$ of trigger leg

Table 4.2: Electron candidate selection requirements on conversion partner leg to the trigger.

are required to have $|\eta_{trk}| < 1.2$ by the CES clustering algorithm, which we describe in greater detail in Section 4.3.2. Table 4.2 summarizes the selection criteria for the partner leg.

We do not require electrons to be matched to a jet in this sample because many of the reconstructed jets are simply the electron showers associated with the conversion legs, and so these events are dissimilar from the jet-embedded electron topology characteristic of HF electrons. Although π^0 decay can result in conversion electrons within jets, direct photon production will result in isolated conversion electrons. Therefore, we do not expect the conversion electrons in this sample to be produced exclusively within a jet. The fraction of the jet energy that is electromagnetic for the conversion electron sample - shown in Figure 4.5 - strongly suggests this conclusion.

For conversion identification, we use the conversion identification variables $\Delta \cot(\theta)$ and sep (see Section 3.2.1 for their application to high E_T electrons) in conjunction with the signed radius of conversion and the number of missing silicon layers. The con-

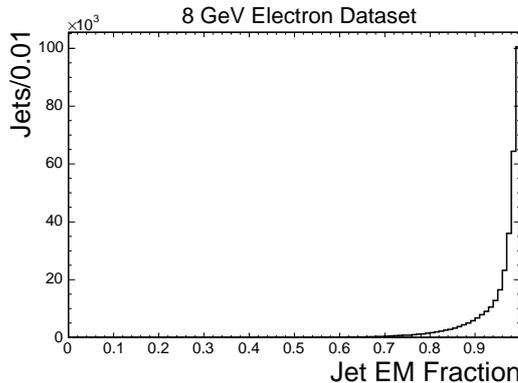


Figure 4.5: Electromagnetic fraction of jets matched to the conversion electron partner leg.

version identification criteria are summarized in Table 4.3. The definition of $\Delta \cot(\theta)$ is simply the difference in $\cot(\theta)$ between the two tracks. The variable sep is defined as the shortest distance between the two tracks in the $r-\phi$ plane at the point that they are parallel. sep has a positive sign when the tracks do not cross in the $r-\phi$ plane; it has a negative sign otherwise. Similarly, the radius of conversion, R_{conv} , is the distance from $(0,0)$ in the $r-\phi$ plane to the same point. We give R_{conv} a negative sign if it is located ‘behind’ $(0,0)$. More precisely, the sign of R_{conv} is given by the dot product of the reconstructed conversion photon momentum vector and the position vector with respect to $(0,0)$ in the $r-\phi$ plane. An illustration of this is shown in Figure 4.6.

The missing silicon layer variable is constructed by extrapolating the track through the silicon and determining whether or not that silicon element is expecting a hit. If both the axial and stereo sides of a silicon layer are expecting at least one hit and none are found on either side, then that layer is considered missing. Counting the total number of missing layers creates the variable. In conjunction with our discussion

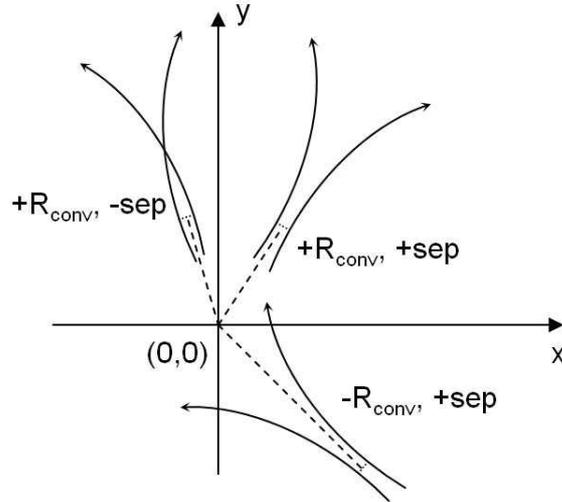


Figure 4.6: Illustration of conversion pairs with positive and negative R_{conv} and positive and negative sep .

Conversion Reconstruction Criteria

$$|\Delta \cot(\theta)| < 0.01$$

$$|sep| < 0.1 \text{ cm}$$

$$\text{Signed } R_{conv} > 6 \text{ cm}$$

At least 3 missing silicon layers

Table 4.3: Electron candidate selection requirements on conversion partner leg to the trigger.

on conversions, this variable is described in greater detail in Section 5.3.

More than 380,000 conversion electrons are found using the selection described here. Although the beamline corrected impact parameter requirement suppresses conversions in this sample, we have applied it for consistency with our general soft electron selection criteria. This requirement affects the radius of conversion by suppressing conversions at high R_{conv} . The missing silicon requirement suppresses conversions at low R_{conv} . By fitting the $\Delta \cot(\theta)$ distribution to the sum of two Gaussians and a

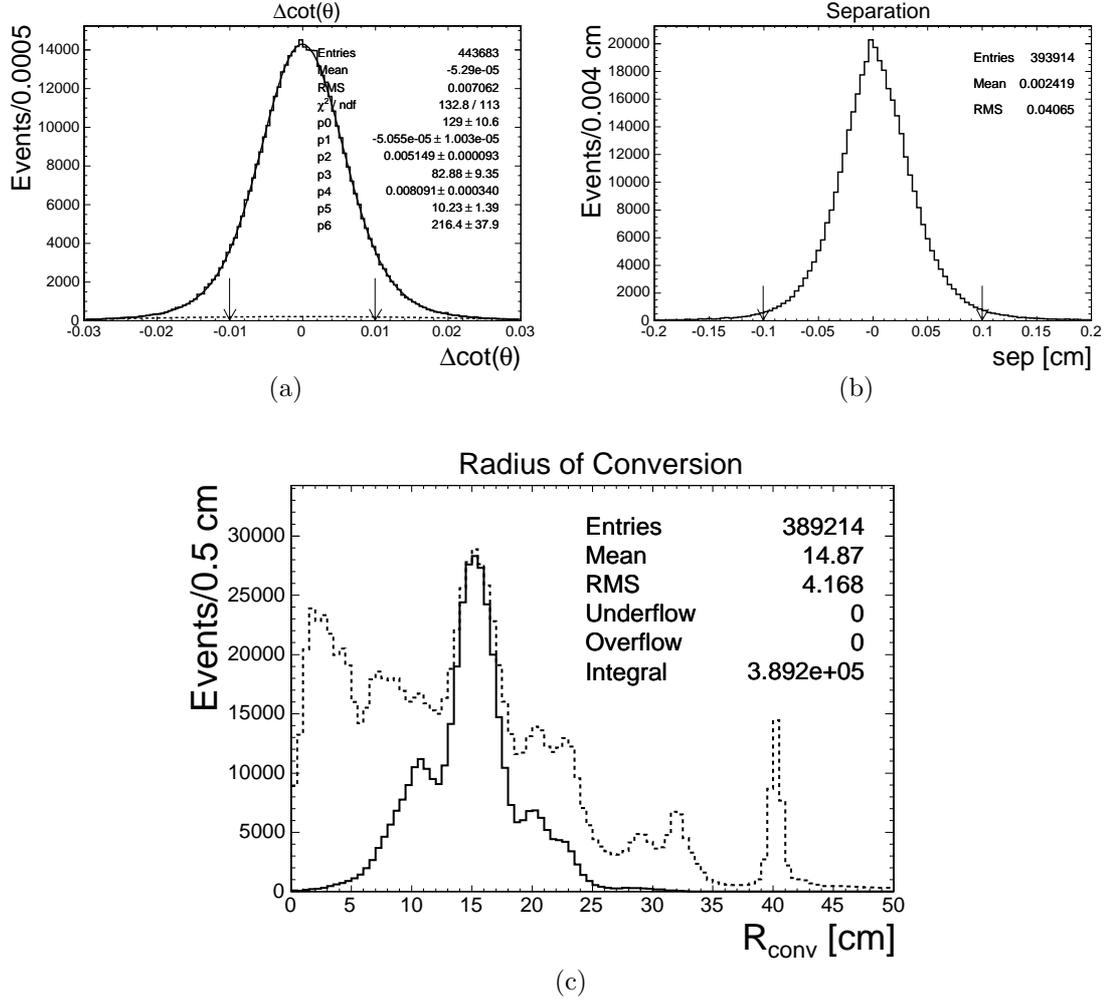


Figure 4.7: Conversion identification distributions, $\Delta \cot(\theta)$ (a), separation (b), and R_{conv} (c) in the 8 GeV electron dataset. Requirements are released on the given variables only to show the tails. $\Delta \cot(\theta)$ is fit to the sum of two Gaussians and a quadratic. The fit is shown in the solid line, while the quadratic component (which we associate with the combinatorial backgrounds) is shown in the dotted line. R_{conv} of the conversion electron pair with (solid line) and without (dotted line) the impact parameter and missing silicon requirements. The high R_{conv} conversions are suppressed due to the impact parameter requirement. The low R_{conv} conversions are suppressed due to the missing silicon layer requirement.

quadratic and associating the quadratic function with the background, we estimate that the conversion sample is more than 98% pure. Figure 4.7 shows the variables

$\Delta \cot(\theta)$, sep , and R_{conv} for the conversion electrons selected. The R_{conv} distribution with the impact parameter and missing silicon layer requirements released is also shown normalized to compare the shape. For reference, the beampipe is located at 1.2 cm, and the COT inner cylinder is at 40 cm.

4.2.2 Generic Track Selection: Fake Electrons

We model fake electrons with generic tracks, collected from events triggered on a 50 GeV jet. The kinematic and environmental similarities to fake electrons in the lepton+jets dataset or other high p_T samples make this sample an appropriate choice. The drawback of using generic tracks is the contamination by real electrons in the sample. Such electrons can come from conversions, Dalitz decays of π^0 , heavy flavor decay, and other sources. For the creation of an SLT_e likelihood template, this contamination has a small effect since we expect that only small fraction of generic tracks are real electrons. However, if we wish to measure the tagging efficiency for fake electrons, we must estimate the contamination. This is done in Chapter 7 when we have the appropriate machinery.

The selection of generic tracks shares the same fiduciality and kinematic requirements as the conversion electron selection. We again select tracks with $p_T > 2 \text{ GeV}/c$ and $|d'_0| < 0.3 \text{ cm}$ that are well measured, extrapolate to the CES and seed a track-based cluster there. However, instead of finding a conversion partner, we suppress conversions for the generic track selection. We use a conversion filter which rejects candidate tracks if another oppositely charged track in the event forms a conversion with $|\Delta \cot(\theta)| < 0.03$, $|sep| < 0.2 \text{ cm}$, and a positively signed conversion radius.

Tracks with more than three missing silicon layers are also rejected as conversions. These requirements are looser than the ones we apply to the conversion electron sample because we want to suppress conversion electrons rather than create a pure sample of them.

Tracks in this sample are required to be within a cone of 0.4 to a fiducial jet¹ with a corrected E_T above 20 GeV. To avoid a bias from the trigger, we reject tracks in the jet that matches to a cluster above the level 2 trigger threshold (40 GeV) unless another such jet exists in the event. The generic track selection is summarized in Table 4.4. We have chosen the generic track selection to be identical to our expected HF electron selection, other than the trigger requirement.

Shown in Figure 4.8 are some of the track parameters of the conversion electrons and generic jets, shown together. The conversion selection we use sculpts the $q d'0$, η , and z_0 of the conversion electron sample. For instance, the three barrel structure of the silicon can be seen in the z_0 distribution. This effect sculpts the η distribution as well because η and z_0 are correlated due to the calorimeter and CES fiduciality requirement. This fiduciality requirement also reduces the number of candidate tracks in both samples in the central $|\eta| < 0.3$ region.

4.3 Electron Identification Variables

In this section we consider the variables used in the soft electron tagger. We study the behavior of the calorimeter and CES using conversions to model HF electrons and

¹In general for this analysis, jets are clustered using the JetClu0.4 algorithm. Jet energies are corrected for detector effects, variations in η , multiple interactions, and the absolute energy scale. See Section 3.3.1 for more details.

Selection Criteria for Generic Tracks
Fiducial to CES Seeds track-based CES cluster Axial segments ≥ 2 Stereo segments ≥ 2 $ z_0 < 60$ cm $p_T > 2$ GeV/c $ d'_0 < 3$ mm
Matched to a jet: ΔR between track and jet < 0.4 Jet $E_T > 20$ GeV (corrected) Jet $ \eta_D < 2.0$
Missing SI layers ≤ 3 No other oppositely-charged track in event with: $ \Delta \cot(\theta) < 0.03$, $ sep < 0.2$ cm, and Signed $R_{conv} > 0$ cm
Trigger Requirement: Jet matched to a Level 2 trigger cluster with $E_T > 40$ GeV No tracks matched to ‘trigger’ jet, unless ≥ 2 jets pass trigger

Table 4.4: The selection criteria for generic tracks.

generic tracks to model fake electrons.

4.3.1 Calorimeter Variables

We use the canonical electron identification variables E_{EM}/p and E_{Had}/E_{EM} to determine whether or not the electromagnetic shower is consistent with an electron in a jet. E_{EM} and E_{Had} are the total electromagnetic and hadronic energy depositions in the two-tower calorimeter cluster (defined below), respectively, and p is the momentum of the track. Limiting the number of calorimeter towers in the cluster to two minimizes the effects of the jet on the electron measurement.

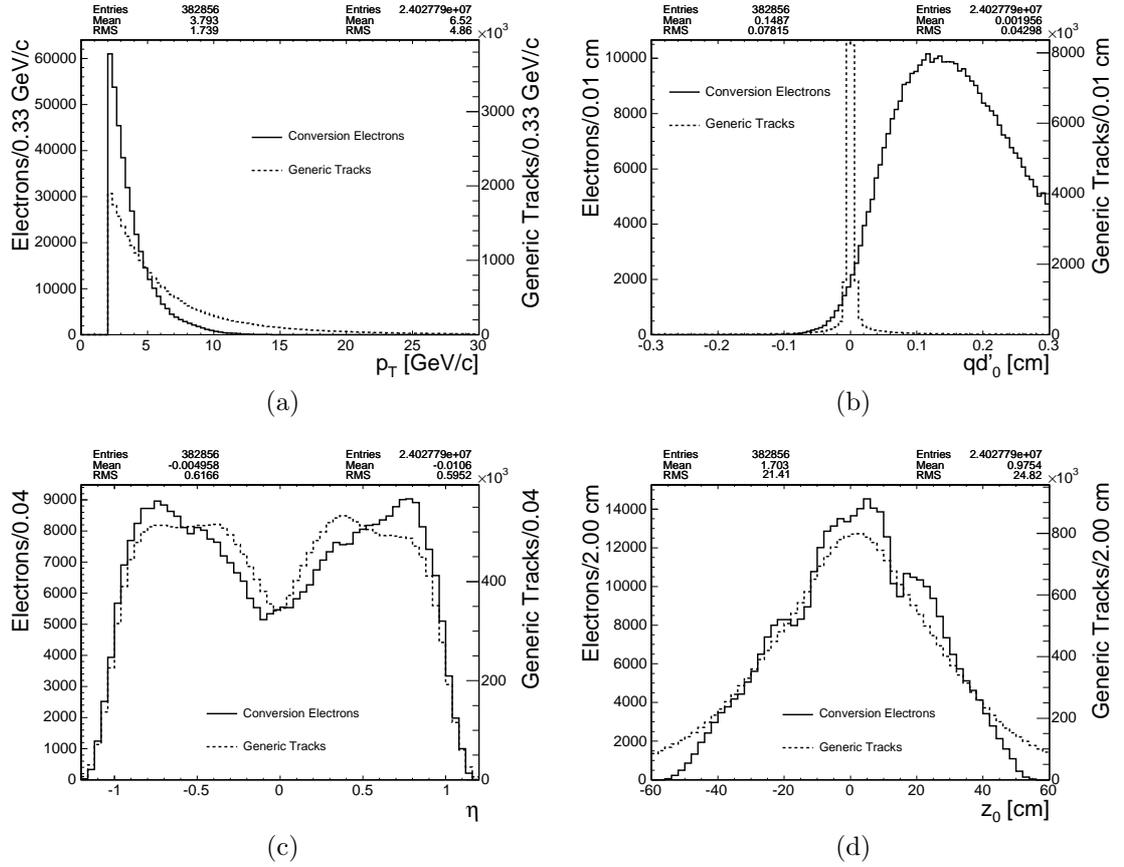


Figure 4.8: Track parameters for the conversion electron (solid line) and generic track (dotted line) samples. The scale and statistics for each are on the left and right, respectively. Parameters shown are p_T (a), qd'_0 (b), η_{trk} (c), and z_0 (d).

The clustering algorithm extrapolates the candidate track to the calorimeter, and the closest calorimeter tower in η and ϕ is chosen as a seed. We combine the seed tower with an adjacent tower in η -space to capture additional energy that may have been deposited outside of the initial seed tower. This adjacent tower is chosen to be the one closer to the track extrapolation in z . In the case that the adjacent tower crosses the $\eta_{det} = 0$ crack, it is not used. For consideration in the cluster, towers are required to have a minimum of 0.1 GeV total E_T (vertex calculated from $z = 0$ cm).

The minimum cluster energy therefore is 0.1 GeV since, occasionally, only one tower is used.

In order to determine the effect of the surrounding jet on the variables E_{EM}/p and E_{Had}/E_{EM} , we construct a track-based isolation variable, I :

$$I \equiv \frac{\sum_{clst} p_T}{p_T}. \quad (4.1)$$

The numerator of this expression is the sum of the transverse momentum of tracks that extrapolate to one of the two towers and has $p_T > 0.5$ GeV/ c . Tracks must also be within $\Delta R \equiv \sqrt{\Delta\eta_{trk}^2 + \Delta\phi^2} \leq 0.4$ of the candidate track. Since only tracks that point to the cluster are considered, the effect of the ΔR requirement is minimal and merely rejects tracks from a different vertex. The denominator is the candidate track's p_T . A track with no other tracks with $p_T > 0.5$ GeV/ c in a cone of 0.4 that extrapolate to the towers, has isolation, by definition, equal to 1. This variable is intended to estimate the degree to which the local environment might affect the electron identification variables used throughout this note. The standard calorimeter-based isolation variable does not suit this purpose. Figure 4.9 shows the isolation distribution for the sample of conversion electrons and generic tracks. Conversion electrons are more isolated since we have not required them to be inside of a jet. Additionally, because the legs must point to different electromagnetic clusters, the partner leg cannot affect the isolation variable.

Figure 4.10 shows the E_{EM}/p distribution for both isolated ($I = 1$) and non-isolated ($I > 1$) conversion electrons and generic tracks. In the isolated case, conversion electrons exhibit a characteristic Gaussian peak at $E_{EM}/p = 1.0$ with a long forward bremsstrahlung tail. In the sample of generic tracks, the E_{EM}/p peaks earlier

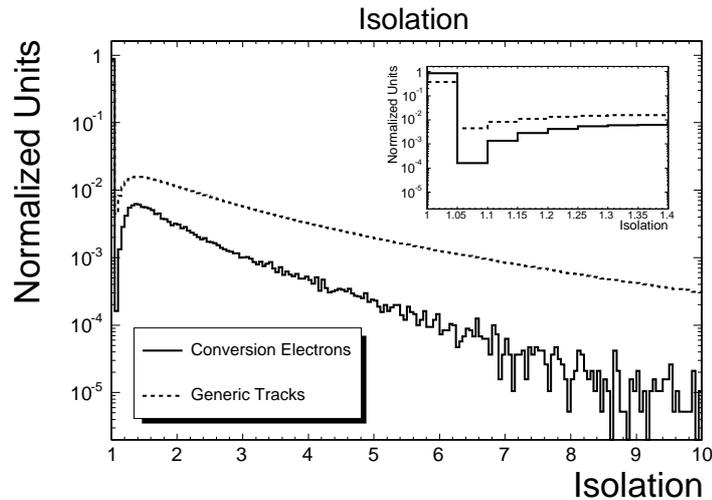


Figure 4.9: The isolation variable I distribution for the sample of conversion electrons and generic tracks. The embedded histogram shows the $I < 1.4$ region. The histograms have been normalized to unit area. As expected, conversion electrons are more isolated than generic tracks.

than for similarly isolated electrons. However, as seen in Figure 4.10, non-isolated electrons have a peak that is shifted and is much broader. The degree to which the peak is broadened and shifted depends on how non-isolated the electron is. A similar effect can be seen in generic tracks.

Figure 4.11 shows the E_{Had}/E_{EM} distribution for both isolated and non-isolated conversion electrons and generic tracks. Electrons peak strongly at 0, whether isolated or non-isolated, although the hadronic energy component is larger for non-isolated tracks. Generic tracks have an increasing E_{Had} contribution relative to the E_{EM} as I increases.

We expect a strong dependence of E_{EM}/p and E_{Had}/E_{EM} on the isolation since the calorimeter towers are coarsely segmented: each tower subtends approximately

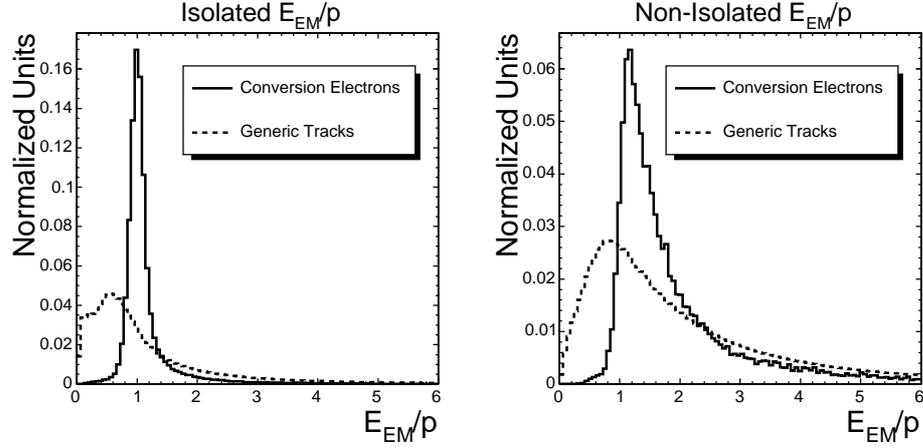


Figure 4.10: E_{EM}/p of electrons from the conversion sample and generic tracks normalized to unit area. Left: Isolated ($I = 1.0$). Right: Non-Isolated ($I > 1.0$). No E_{Had}/E_{EM} requirement is made.

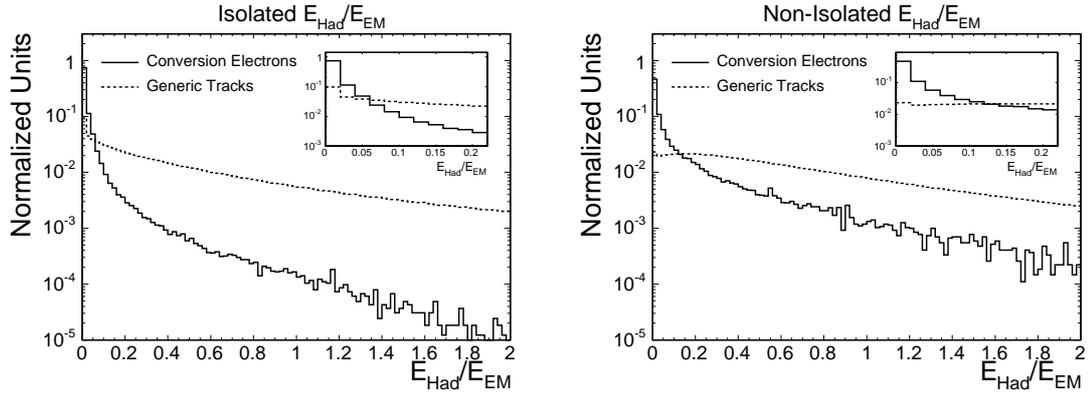


Figure 4.11: E_{Had}/E_{EM} of electrons from the conversion sample and generic tracks normalized to unit area. Left: Isolated ($I = 1.0$). Right: Non-Isolated ($I > 1.0$). The embedded histograms show the $E_{Had}/E_{EM} < 0.22$ region. No E_{EM}/E_p requirement is made.

15° in ϕ -space and approximately 0.1 in η -space. As we noted when constructing the conversion electron sample in Section 4.2.1, conversion electrons are often not embedded in a jet and so have a substantially different topology than what we might

expect from HF electrons. Therefore, we turn to MC to examine how HF electrons behave. Figure 4.12 show the E_{EM}/p and E_{Had}/E_{EM} distributions for HF electrons and fake electrons in $t\bar{t}$ MC. Candidate tracks are selected with the same criteria (except for the trigger) as generic tracks (Table 4.4). The HF electron distributions are clearly affected by the environment, exhibiting a broad, shifted E_{EM}/p peak, and a slowly declining E_{Had}/E_{EM} distribution.

Based on the distributions from $t\bar{t}$ MC shown in Figure 4.12, we require that

- $0.6 < E_{EM}/p < 2.5$
- $E_{Had}/E_{EM} < 0.2$

for an SLT_e tag. We address in Section 6.4 how well-modeled these requirements are in MC. Table 4.5 provides the efficiencies for each requirement for the conversion electron sample, generic tracks, and different types of electrons in the $t\bar{t}$ MC sample. A couple of points are worth noting. First, the efficiency for the calorimeter requirements for the conversion electrons in our pure data sample and conversion electrons in $t\bar{t}$ MC are different because the samples were constructed differently. The data sample requires that the two electron legs not share an electromagnetic cluster in the calorimeter, whereas the MC sample has no such requirement. When both legs are present in a single cluster, this can distort the expected E_{EM}/p distribution. Second, the jet environment is much greater in the MC datasets, reducing the overall efficiency. We also note that generic tracks in data have a small electron contamination content which makes the calorimeter requirements more efficient than for pure fake electrons in MC.

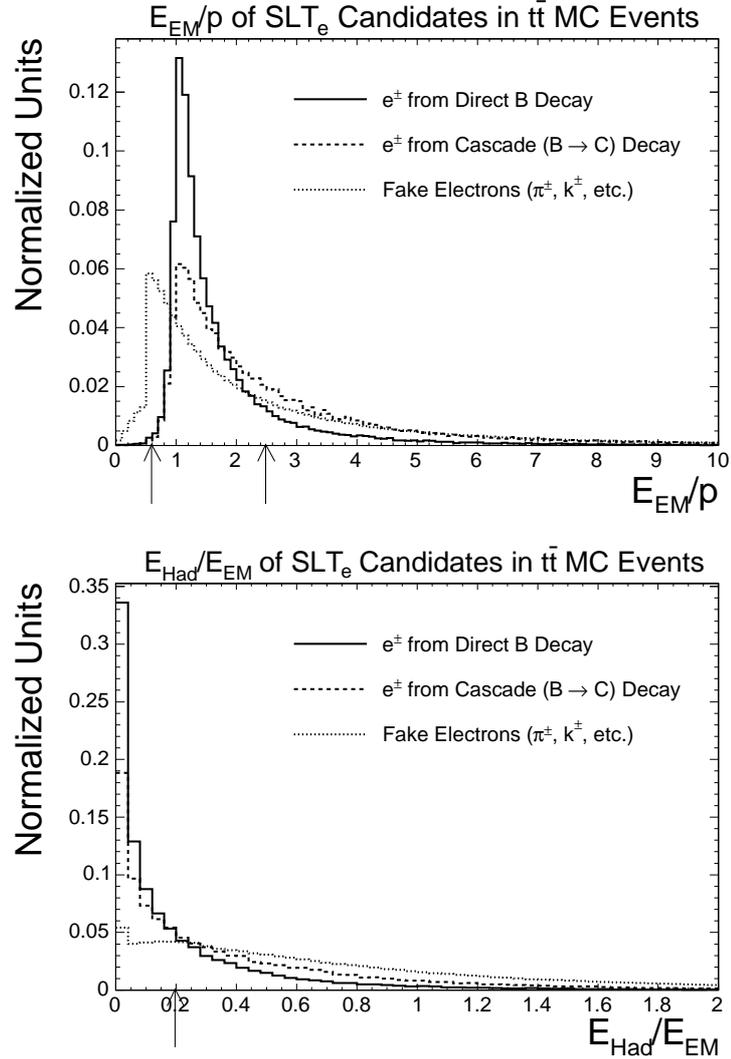


Figure 4.12: E_{EM}/p (top) and E_{Had}/E_{EM} (bottom) of the calorimeter clusters of electrons from b decay and cascade decay and fake electrons in $t\bar{t}$ MC events.

4.3.2 Shower Max Clustering

The CES clustering algorithm begins by extrapolating the candidate SLT_e tracks to the CES. The closest wire and strip to the extrapolated position is used as a seed to form a cluster. The cluster is then formed by including the closest 7 wires/strips

Sample	$E_{EM}/p \epsilon$	$E_{Had}/E_{EM} \epsilon$	Combined ϵ
e^\pm from <i>two-tower</i> conversions (data)	87%	94%	82%
e^\pm from conversions ($t\bar{t}$ MC)	43%	41%	23%
e^\pm from b -jets ($t\bar{t}$ MC)	85%	68%	61%
e^\pm from c -jets ($t\bar{t}$ MC)	64%	47%	35%
Generic tracks (data)	51%	26%	14%
Fake electrons ($t\bar{t}$ MC)	47%	16%	8%

Table 4.5: Efficiencies of the $0.6 < E_{EM}/p < 2.5$ requirement and $E_{Had}/E_{EM} < 0.2$ requirement on different samples. Statistical errors are $< 1\%$ absolute.

around and including the seed. The small window is preferred to prevent nearby tracks from affecting the electron identification variables. Wires/strips are only included in the cluster if they are above a threshold of 60/150 MeV (uncorrected). Initially for cluster formation, one wire and one strip above threshold are required to form a cluster.

We can create a few variables to distinguish between electrons and hadrons. For our study of the CES variables, we first apply the calorimeter requirements ($0.6 < E_{EM}/p < 2.5$ and $E_{Had}/E_{EM} < 0.2$) from the previous section, but we note that this requirement will sculpt the CES distributions to look more “electron-like.”

Figures 4.13 and 4.14 show the dependence of both the number of strips and wires in a cluster as a function of p_T and η after the calorimeter requirements have been applied. There is a strong dependence on p_T due to the greater amount of energy deposited in the CES with increasing transverse momentum. The η dependence has several causes. Tracks with higher η but the same p_T have higher momentum, resulting in greater energetic deposition. Additionally, there is increasing detector material traversed by the electron at higher η resulting in an earlier shower, and

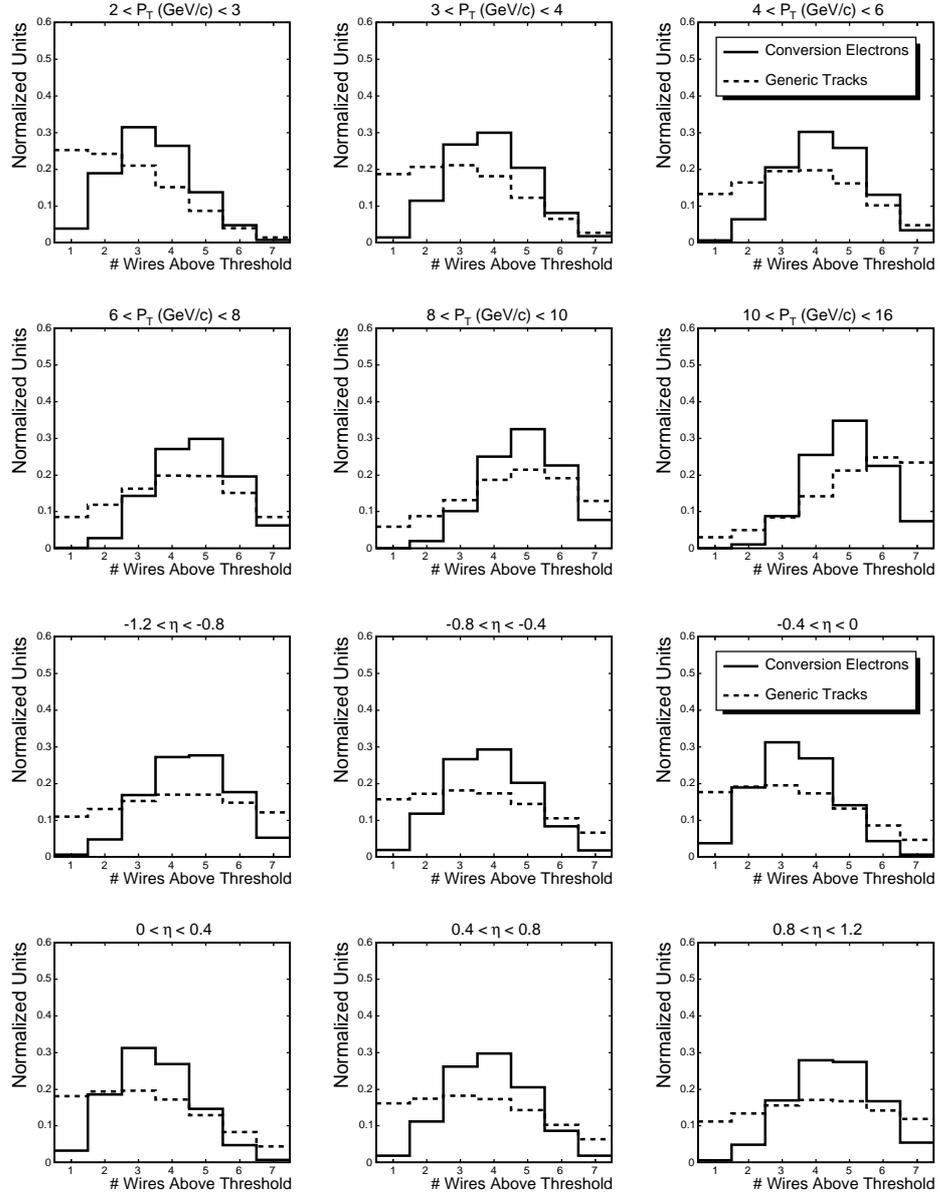


Figure 4.13: Number of wires in a CES cluster as a function of p_T (first two rows) and η (last two rows) for electrons (solid line) and tracks from jets (dotted line). Histograms have been normalized to unit area.

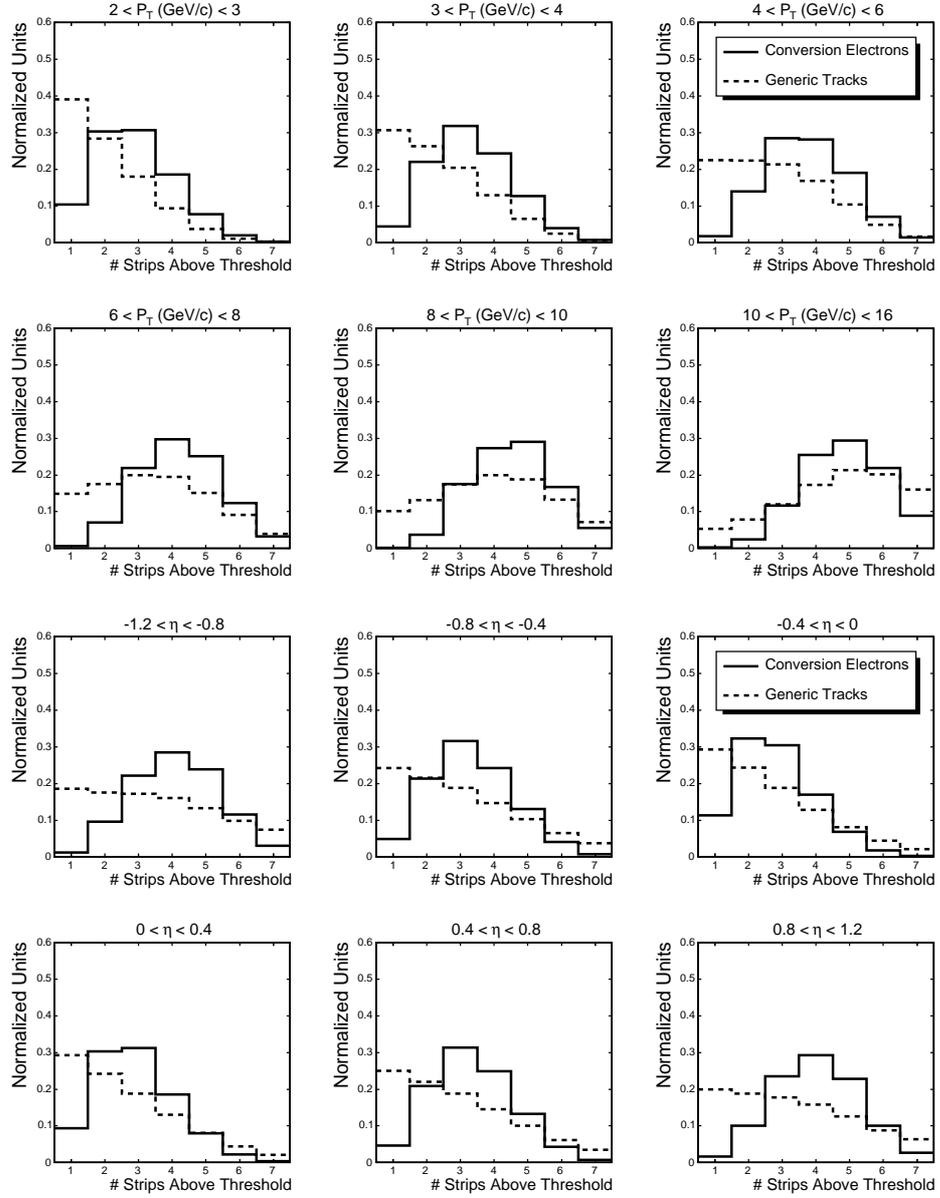


Figure 4.14: Number of strips in a CES cluster as a function of p_T (first two rows) and η (last two rows) for electrons (solid line) and tracks from jets (dotted line). E_{EM}/p and E_{Had}/E_{EM} requirements have been applied. Histograms have been normalized to unit area.

electrons at high η do not intersect the wires or strips transverse to the plane. The wire/strip thresholds used have not been corrected for variations in gain when they are considered for the cluster (they are corrected for the higher level electron identification variables we consider later, such as Δ and χ^2). This effect introduces a slight asymmetry.

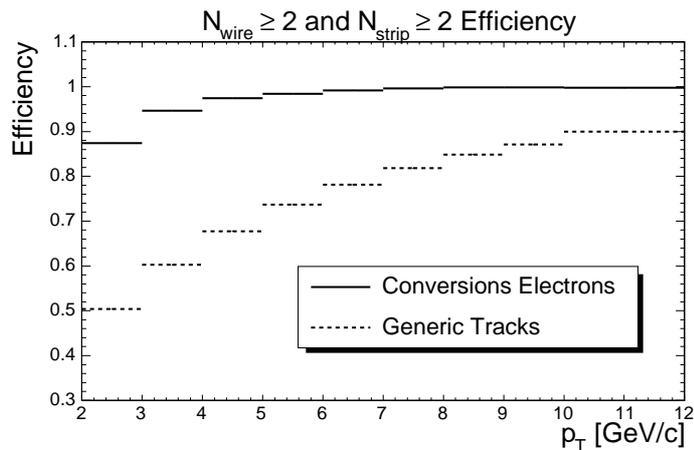


Figure 4.15: Combined efficiency of the $N_{wire} \geq 2$ and $N_{strip} \geq 2$ requirements on electrons and generic tracks as a function of p_T . Calorimeter requirements have been applied. Note the zero suppression.

In order to tag the track, we require that a minimum of 2 strips and 2 wires are above threshold in the CES cluster. This requirement is very effective at removing low p_T hadrons which deposit very little energy in the CES and calorimeter. Figure 4.15 shows the efficiency as a function of p_T for both conversion electrons and generic tracks. The overall sensitivity to the environment of the number of wires and strips in the cluster is small.

4.3.3 χ^2 Variables

We can also calculate the energy and position of the cluster:

$$E = \sum_{i=1}^N E_i; \quad X = \sum_{i=1}^N \frac{X_i E_i}{E} \quad (4.2)$$

where i runs over all N of the channels above threshold (up to 7) included in the cluster. The transverse profile of the cluster is also compared to a 10 GeV test beam through a χ^2 “goodness-of-fit” test:

$$\chi^2 = \frac{1}{4} \sum_{i=1}^N \frac{y_i^2 - \hat{y}_i^2}{\hat{\sigma}_i^2} \quad (4.3)$$

where y_i is the measured energy fraction (E_i/E) for channel i , and \hat{y}_i and $\hat{\sigma}_i$ are the fraction and RMS of the test beam profile. We scale each χ^2 by its momentum according to the formula:

$$\chi^2 \rightarrow 0.1792 \cdot \chi^2 \cdot 2.11^{\ln(p)} \quad (4.4)$$

because the test beam was performed at 10 GeV, and this scaling was found to remove the dependence on the electron energy.

Figure 4.16 shows the χ^2 variables for conversion electrons and generic tracks after E_{EM}/p , E_{Had}/E_{EM} , and wire/strip requirements have been applied. A singularity in the χ^2 distribution (see Figure 4.17) occurs when only one wire is used to calculate the χ^2 value. Since the energy profile is normalized in Equation 4.3, a cluster with a single wire or strip has precisely the same χ^2 value regardless of the amount of energy deposited. The singularity gets smeared out by the scaling in Equation 4.4. The same effect is seen in the strips view.

The χ^2 variable is a well discriminating variable between real electrons and fake electrons. Although the χ^2 of conversion electrons is stable across p_T , generic tracks

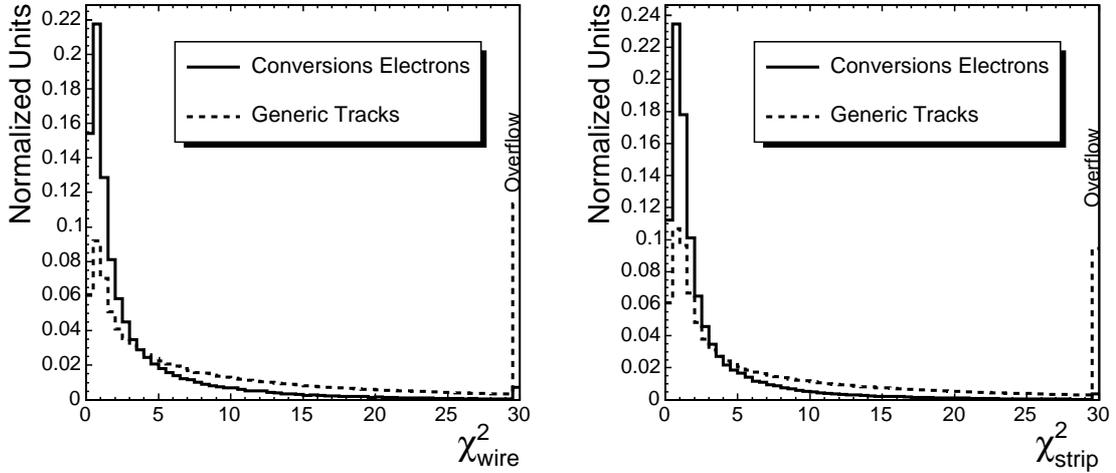


Figure 4.16: χ^2 in wire (left) and strip (right) views, normalized to unit area. Electrons are shown in solid lines while generic tracks are shown in dotted lines. The histograms have been normalized to unit area, and the last bin is an overflow. Calorimeter and wire/strip requirements have been made.

show a p_T dependence, evincing greater separation at high p_T and a longer tail in both the wire and strip views. Figures 4.18 and 4.19 show this variation. The χ^2 variable has a small isolation dependence even for electrons, as shown in Figure 4.20.

4.3.4 Δ_{wire} and Δ_{strip} Variables

The CES variable $q\Delta_{wire}$ is the distance (in cm) between the track extrapolation and the center of energy of the CES wire cluster in the transverse ($r - \phi$) plane signed by the charge of the track (see Equation 4.2). Similarly, Δ_{strip} is the distance between the track extrapolation and the strip cluster in the z plane. We correct the extrapolated track position for the misalignment in the CES.

Again, we first apply the E_{EM}/p , E_{Had}/E_{EM} , and wire/strip requirements to the

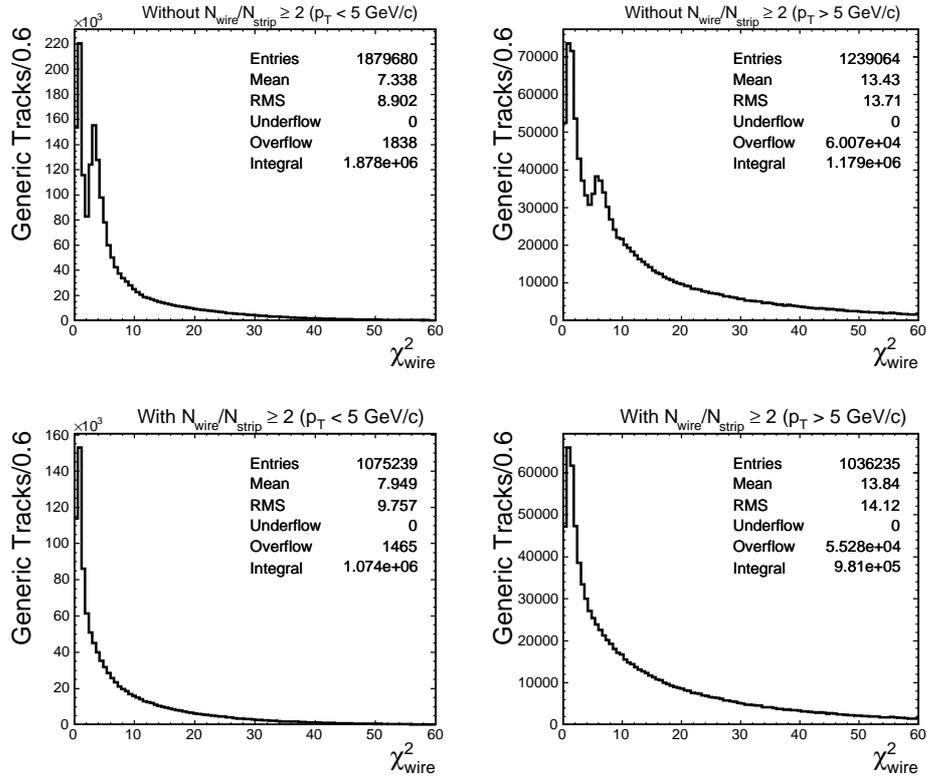


Figure 4.17: χ_{wire}^2 for generic tracks. Calorimeter requirements only have been applied. Rows are divided between having the wire/strip requirement applied (bottom) and not applied (top). Columns are divided between p_T less than (left) and greater than (right) 5 GeV/c. The second peak disappears when the 2 or more wires and 2 or more strips are required. The χ_{strip}^2 distributions are similar.

conversion electron and generic track samples. Both $q\Delta_{wire}$ and Δ_{strip} are roughly Gaussian with means centered at zero for both conversion electrons and generic tracks, but the width varies as a function of p_T . In Figures 4.21 and 4.22, we fit the $q\Delta_{wire}$ and Δ_{strip} distributions to the sum of two Gaussians (with fixed means) in bins of p_T , and parameterize the narrower width according to the multiple scattering function

$$\sigma(p_T) = a + b/p_T. \quad (4.5)$$

For the electron distributions, we can use the fits to re-normalize the distribution

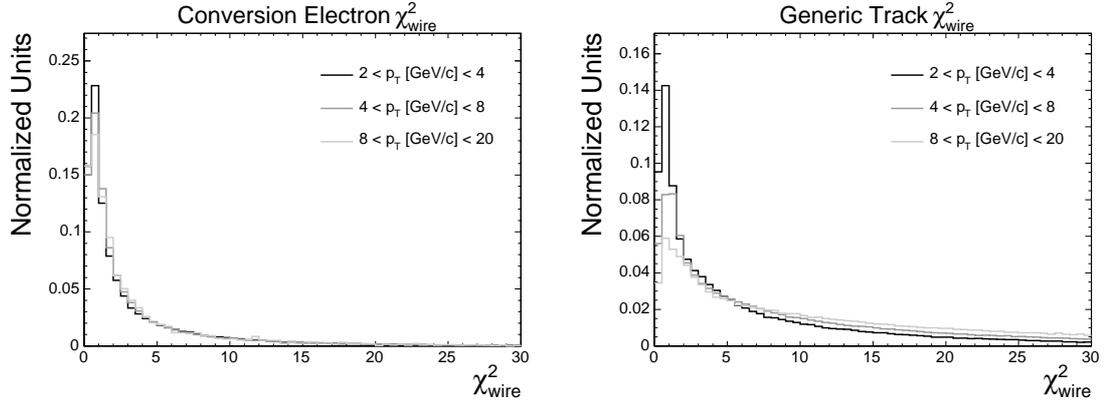


Figure 4.18: χ_{wire}^2 for conversion electrons (left) and generic tracks (right) in the wire view as a function of p_T . Electrons exhibit stable behavior over a broad range of transverse momentum, whereas generic tracks exhibit a longer tail at high p_T . All histograms have been normalized to unit area to show the relative shape. Calorimeter and wire/strip requirements have been made.

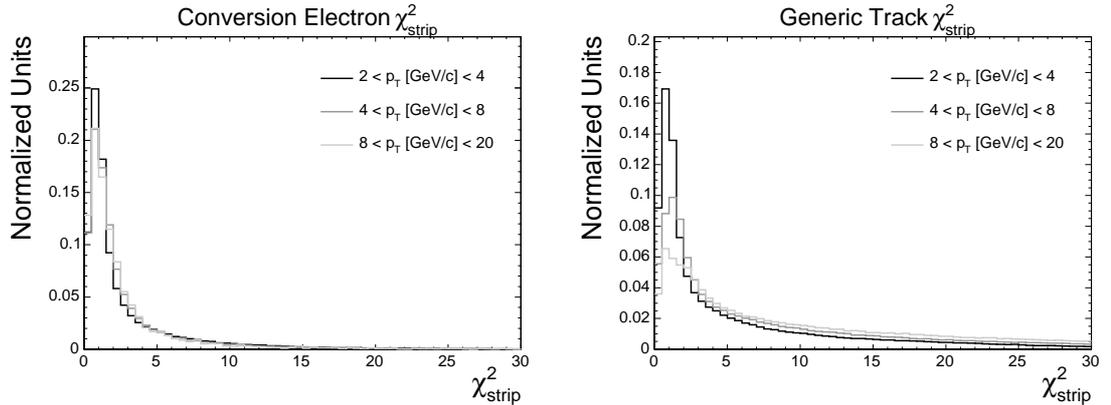


Figure 4.19: χ_{strip}^2 for conversion electrons (left) and generic tracks (right) in the wire view as a function of p_T . Electrons exhibit stable behavior over a broad range of transverse momentum, whereas generic tracks exhibit a longer tail at high p_T . All histograms have been normalized to unit area to show the relative shape. Calorimeter and wire/strip requirements have been made.

as a function of p_T thereby removing the p_T dependence. We use the new variables

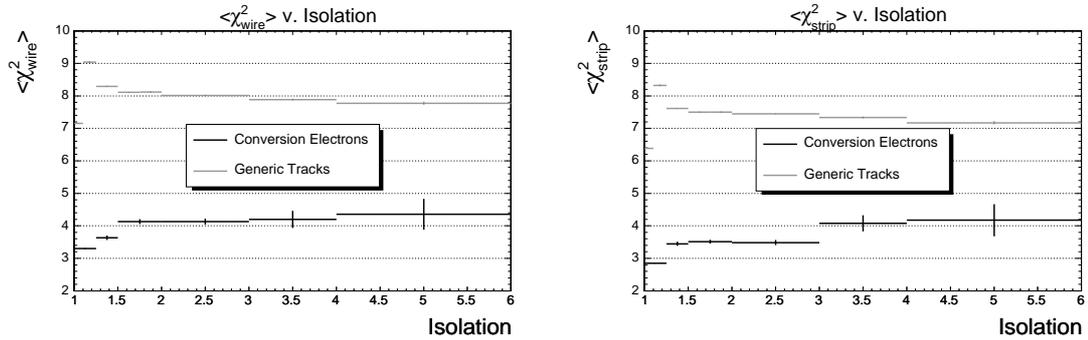


Figure 4.20: Mean χ^2 in the wire (left) and strip (right) view for conversion electrons (black) and generic tracks (gray) as a function of isolation. Calorimeter and wire/strip requirements have been made.

$q\Delta_{wire}/\sigma_{wire}$ and $\Delta_{strip}/\sigma_{strip}$ where

$$\sigma_{wire} = 0.19 + 1.2 \cdot p_T \quad (4.6)$$

and

$$\sigma_{strip} = 0.45 + 0.79 \cdot p_T. \quad (4.7)$$

As a cross-check, we also consider the mean of the normalized distributions as a function of p_T and η . Figure 4.23 shows the mean of $q\Delta_{wire}/\sigma_{wire}$ and $\Delta_{strip}/\sigma_{strip}$ as a function of p_T and η . No discernable trend can be found in $q\Delta_{wire}/\sigma_{wire}$. For $\Delta_{strip}/\sigma_{strip}$, there is a bias as a function of η due to an alignment problem. We note this effect, but do not correct for it.

We look at the dependence of each normalized variable on p_T for both conversion electrons and generic tracks. Because of the normalization, the width of electron distributions no longer is dependent on p_T . However, generic tracks do show a strong dependence, with increasing separation at high p_T . Figure 4.24 shows the dependence in both the wire and strip views. The isolation of the track has a negligible impact

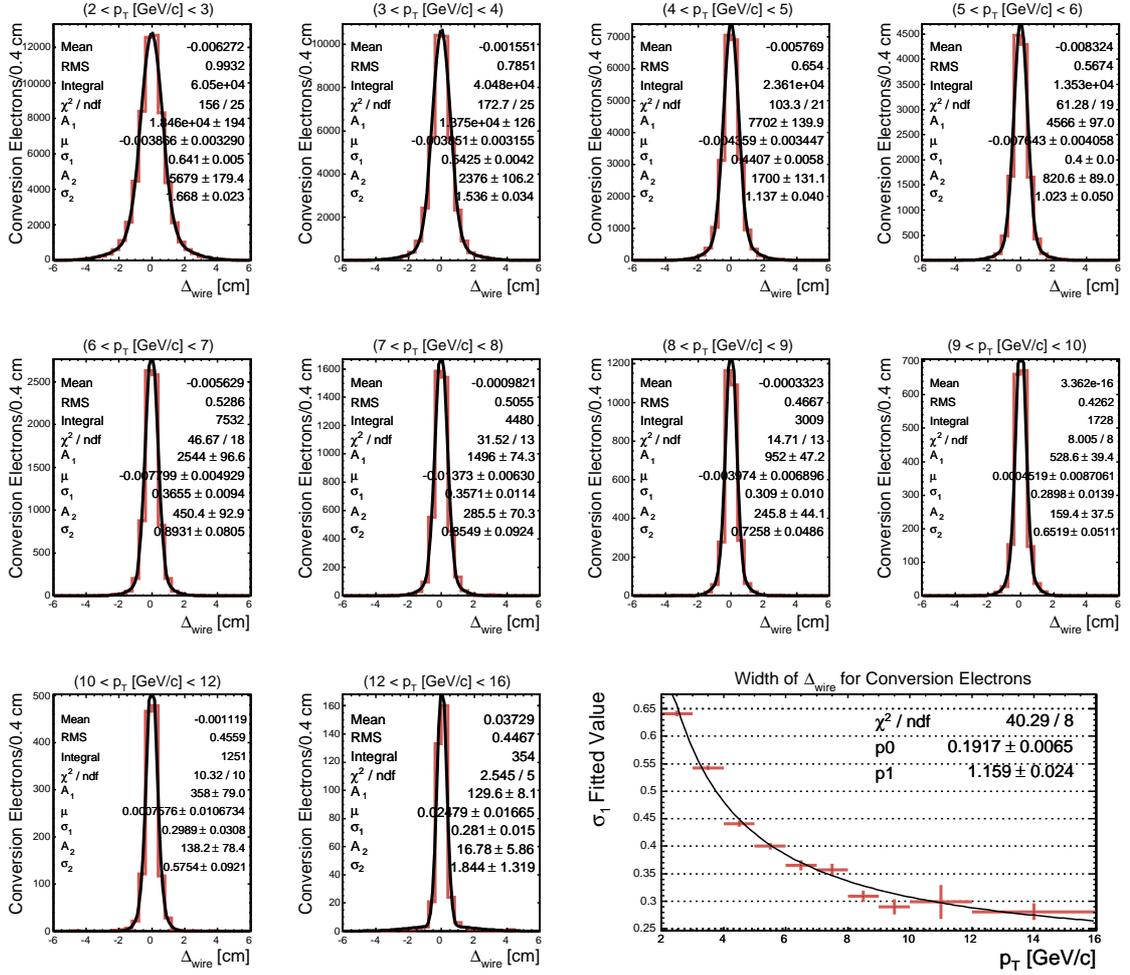


Figure 4.21: $q\Delta_{wire}$ as a function of p_T fit to the sum of two Gaussians constrained to have the same mean, μ . Calorimeter and wire/strip requirements have been made.

on either the mean or the width of the variables.

4.3.5 Likelihood Construction

Using the conversion electron and generic track templates for the tagging variables $\Delta_{wire}/\sigma_{wire}$, $\Delta_{strip}/\sigma_{strip}$, χ_{wire}^2 , and χ_{strip}^2 , we can construct a likelihood algorithm which, combined with the E_{EM}/p , E_{Had}/E_{EM} , and wire/strip requirements, will iden-

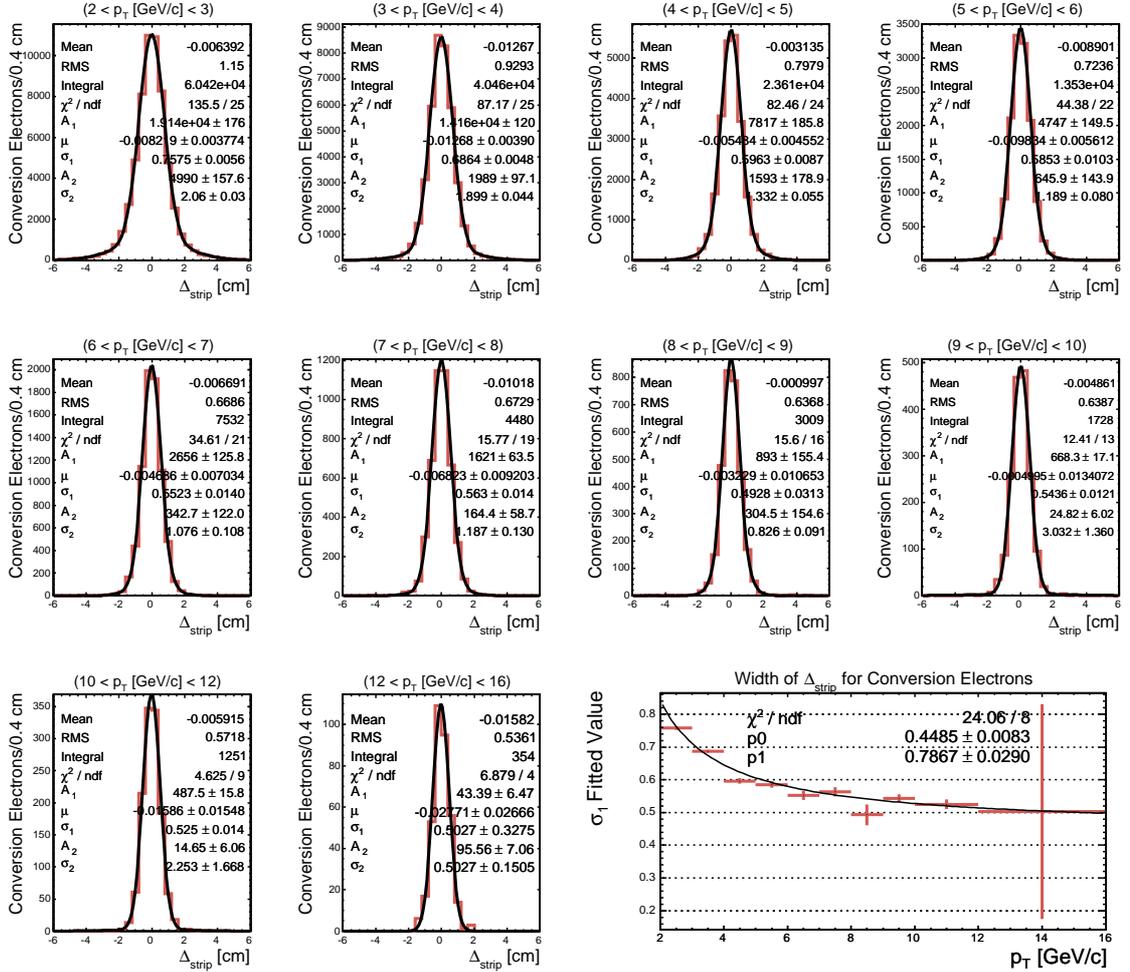


Figure 4.22: $q\Delta_{strip}$ as a function of p_T fit to the sum of two Gaussians constrained to have the same mean, μ . Calorimeter and wire/strip requirements have been made.

tify electrons while suppressing fake electron backgrounds in a high p_T environment.

In order to account for the correlations between the four CES variables, we use a four dimensional histogram as a template with the following binning:

- χ_{wire}^2 and χ_{strip}^2 : [0, 1, 2, 4, 6, 8, 12, 18, 28)
- $|\Delta_{wire}/\sigma_{wire}|$ and $|\Delta_{strip}/\sigma_{strip}|$: [0, 0.8, 1.5, 2.0, 2.5, 3.0, 5.0).

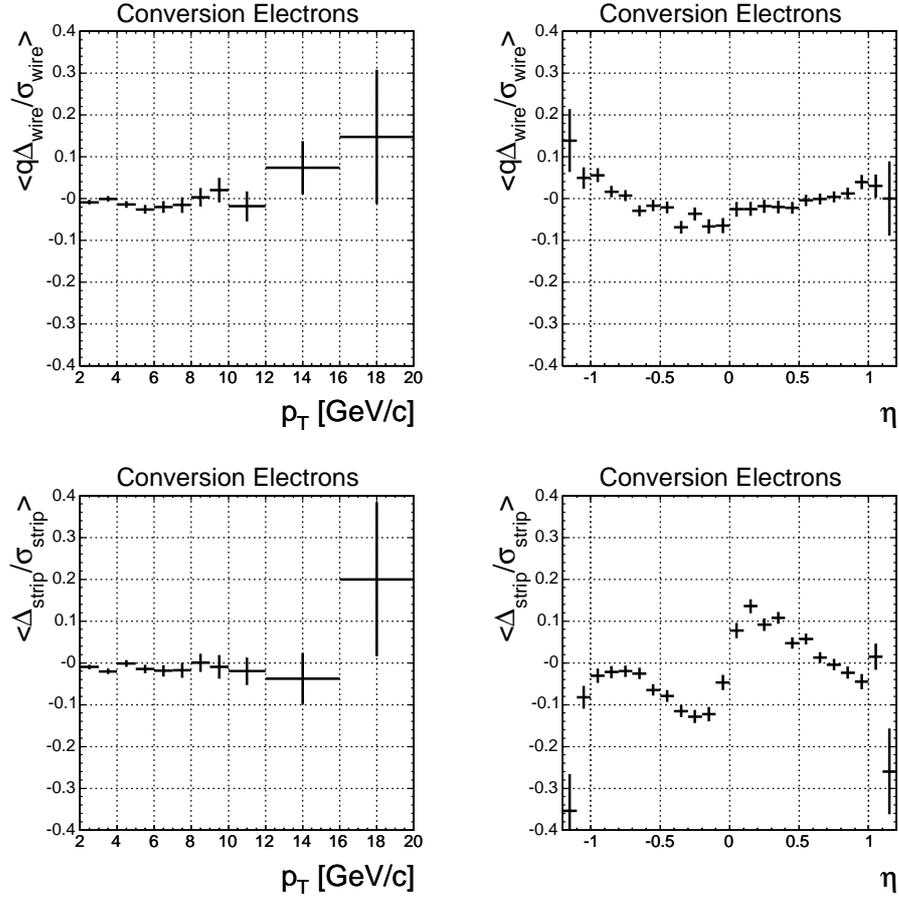


Figure 4.23: Electron $\langle q\Delta_{wire}/\sigma_{wire} \rangle$ (top) and $\langle \Delta_{strip}/\sigma_{strip} \rangle$ (bottom) as a function of p_T (left) and η (right). The errors shown are given by the RMS divided by \sqrt{N} , where N is the number of points. Thus, the large errors at high p_T correspond to fewer data-points, not larger RMS. Calorimeter and wire/strip requirements have been made.

The binning was chosen to maximize the sensitivity in the space of variables where discrimination is most powerful. That is, we chose small bin sizes where the final likelihood is most ambiguous in deciding between electron or fake.

The likelihood is calculated according to the formula

$$\mathcal{L} \equiv \frac{S_i}{S_i + B_i} \quad (4.8)$$

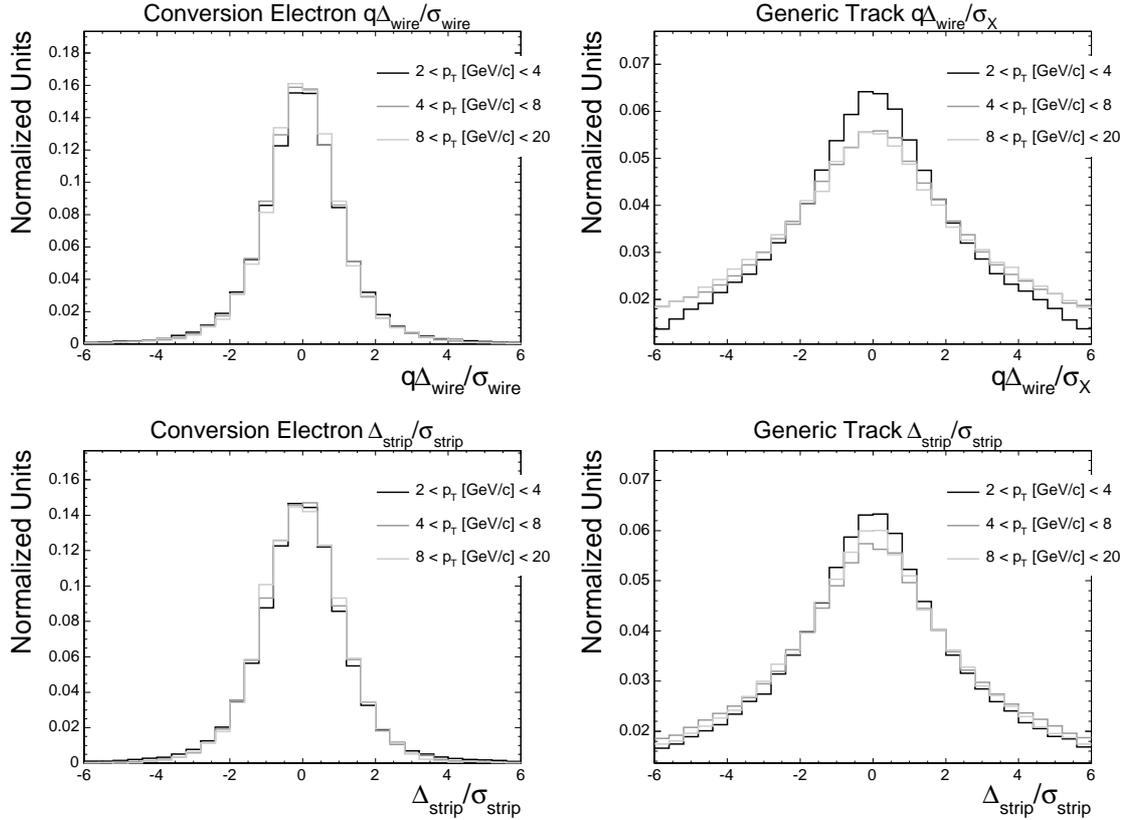


Figure 4.24: Electron (left) and generic tracks (right) distributions of $q\Delta_{\text{wire}}/\sigma_{\text{wire}}$ (top) and $\Delta_{\text{strip}}/\sigma_{\text{strip}}$ (bottom) as a function of p_T . Each histogram has been normalized to unity to show the relative shape. Calorimeter and wire/strip requirements have been applied to both the electron and generic track samples.

where S_i and B_i are the normalized value of the signal and background template histograms in the i^{th} bin. If one of the variables is found to be outside the bounds of the bins, we identify the candidate track as a fake electron by giving it a likelihood equal to 0. Figure 4.25 shows the efficiency if a likelihood selection is made on that value or greater. The maximum likelihood value possible is 0.9 due to the presence of fakes in even the most ‘electron-like’ template bin.

We select 3 different operating points in the tagger implementation: $\mathcal{L} > 0.55$,

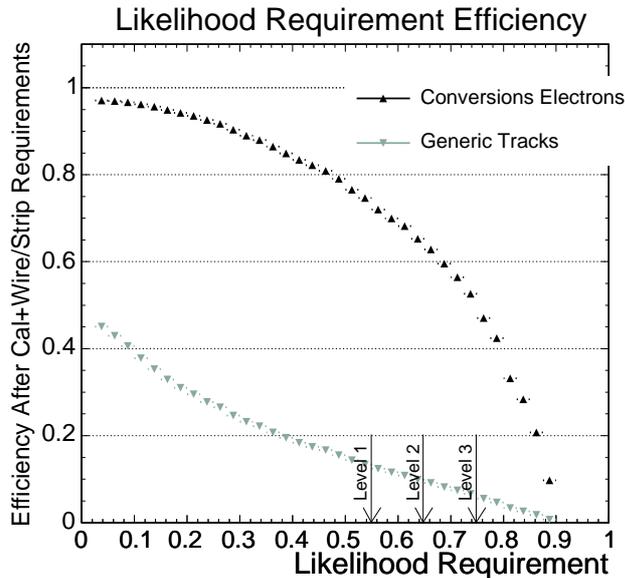


Figure 4.25: Efficiency of a requirement made at that place on the likelihood. This is not the overall efficiency of the tagger since calorimeter and wire/strip requirements have already been applied.

0.65, 0.75. We call these operating points *tag levels* 1, 2, and 3, respectively. With different available operating points, the preferred choice can be determined later by the requirements of the physics analysis.

4.4 Tagger Definition

The SLT_e tagger algorithm consists of selecting tracks with the requirements shown in Table 4.6. In summary, tracks must be well-measured with $p_T > 2 \text{ GeV}/c$, $|d'_0| < 0.3 \text{ mm}$, and $|z_0| < 60 \text{ cm}$. Candidate tracks must be matched to jets within a cone of $\Delta R \leq 0.4$. Tracks are rejected as conversions if a partner track is found or if too many silicon layers are missing. A two-tower calorimeter cluster is formed about

the track, and an energetic deposition consistent with the presence of a single electron is made according to the requirements: $0.6 < E_{EM}/p < 2.5$ and $E_{Had}/E_{EM} < 0.2$. A CES cluster, up to 7 wires and strips, is also formed about the track. At least two wires and two strips must populate the CES cluster or the track is rejected. A likelihood value is calculated using templates of the χ^2 and Δ variables from the conversion electron and generic track samples. If $\mathcal{L} > 0.55, 0.65, 0.75$ - depending on the choice of operating point - the track is considered *tagged*.

For convenience, we define the *taggable* selection as those requirements up to, and including, the E_{Had}/E_{EM} and E_{EM}/p requirements. This is summarized in Table 4.6. The rest of the requirements - the number of wires and strips in the cluster and the \mathcal{L} requirements - define a tag, summarized in Table 4.7. Note that the requirements are similar to those for the generic tracks sample selection (Table 4.4).

In addition to tagging electrons, we can also tag tracks as “not electrons,” also referred to as “anti-electrons.” This will be useful particularly in the following chapter when we wish to construct a sample of prompt tracks. Our procedure is to use reversed calorimeter selection criteria and require two out of four CES variables to be unlike electrons. These requirements are summarized in Table 4.8.

In the conversion sample, only 10 electrons out of nearly 400,000 met these criteria. About 3.5% of the fake sample met the criteria. When using these criteria, it is important to be aware that they preferentially select higher p_T tracks, since fakes are more readily distinguishable from electrons at higher p_T .

Taggable Requirements
Fiducial to CES Seeds track-based CES cluster Axial segments ≥ 2 Stereo segments ≥ 2 $ z_0 < 60$ cm $p_T > 2$ GeV/c $ d'_0 < 3$ mm
Matched to a jet: ΔR between track and jet < 0.4 Jet $E_T > 20$ GeV (corrected) Jet $ \eta_D < 2.0$
Missing SI layers ≤ 3 No other oppositely signed track in event such that: $ \Delta \cot(\theta) < 0.03$ $ sep < 0.2$ cm Signed $R_{conv} > 0$ cm
Two-tower Calorimeter cluster: $0.6 < E_{EM}/p < 2.5$ $E_{Had}/E_{EM} < 0.2$

Table 4.6: The selection criteria to pass taggable requirements.

Tag Requirements
CES cluster: $N_{wires} \geq 2$ $N_{strips} \geq 2$
Likelihood formed from χ^2 and Δ/σ variables: $\mathcal{L} > 0.55, 0.65, 0.75$

Table 4.7: The selection criteria to pass tag requirements.

4.5 Tagging Efficiency

In this section, we study the dependence of the SLT_e on different variables. Figures 4.26, 4.27, and 4.28 show the tagging efficiency as a function of track p_T , η ,

Anti-electron Tag Requirements
Calorimeter Cluster: $E_{EM}/P < 0.5$ $E_{Had}/E_{EM} > 0.3$
Two out of four CES requirements: $\chi_{strip}^2 > 28$ $\chi_{wire}^2 > 28$ $ \Delta_{wire}/\sigma_{wire} > 5$ $ \Delta_{strip}/\sigma_{strip} > 5$

Table 4.8: The selection criteria for anti-electrons.

and isolation for conversion electrons and generic tracks. On the left, the efficiency has been divided up into its composite parts, shown sequentially. The efficiencies are defined as

- Calorimeter efficiency: # of taggable tracks before calorimeter requirements have been imposed divided by # of taggable tracks (including calorimeter requirements).
- Wire+strip efficiency: # of taggable tracks that have ≥ 2 wires and ≥ 2 strips in the CES cluster divided by # of taggable tracks.
- Likelihood efficiency: # of taggable tracks passing wire+strip requirements and likelihood requirements divided by # of taggable tracks passing wire+strip requirements.
- Overall efficiency: # of taggable tracks passing all requirements divided by # of taggable tracks before calorimeter requirements are imposed.

By definition, the overall efficiency is the product of the previous three.

On the right is shown the efficiency of the three tag levels. The average efficiency of the SLT_e in the conversion electron sample at these operating points is 59%, 51%, and 39%, for tag levels 1, 2, and 3, respectively. As shown in the figure, the fake tagging rate varies strongly as a function of p_T . Averaged over the entire generic track sample without correcting for the electron contamination, we find that the average tagging efficiency is 1.1%, 0.8%, and 0.5%, for each tag level.

The declining tagging efficiency as a function of isolation is an unfortunate by-product of fixed calorimeter requirements. Although we could have implemented isolation dependent calorimeter requirements to improve our signal efficiency in the $I \gg 1.0$ region, this would also increase the fake rate substantially. As seen in Figure 4.29, the calorimeter requirement sculpts the isolation distribution substantially.

Another point worth noting is that the isolation variable, by definition, is correlated with p_T . This manifests itself particularly at low values of isolation that are not identically equal to one. In this case, the track p_T must be high to accommodate such a low isolation value. Examination of Equation 4.1 shows why this must be true. This results in a discontinuity across isolation for measurements like the wire/strip efficiency.

Finally, we note that the efficiency is not constant as a function of η and is not even perfectly symmetric about $\eta = 0$. This behavior for the wire/strip requirements was discussed in Section 4.3.2. The calorimeter requirements also exhibit a small asymmetry.

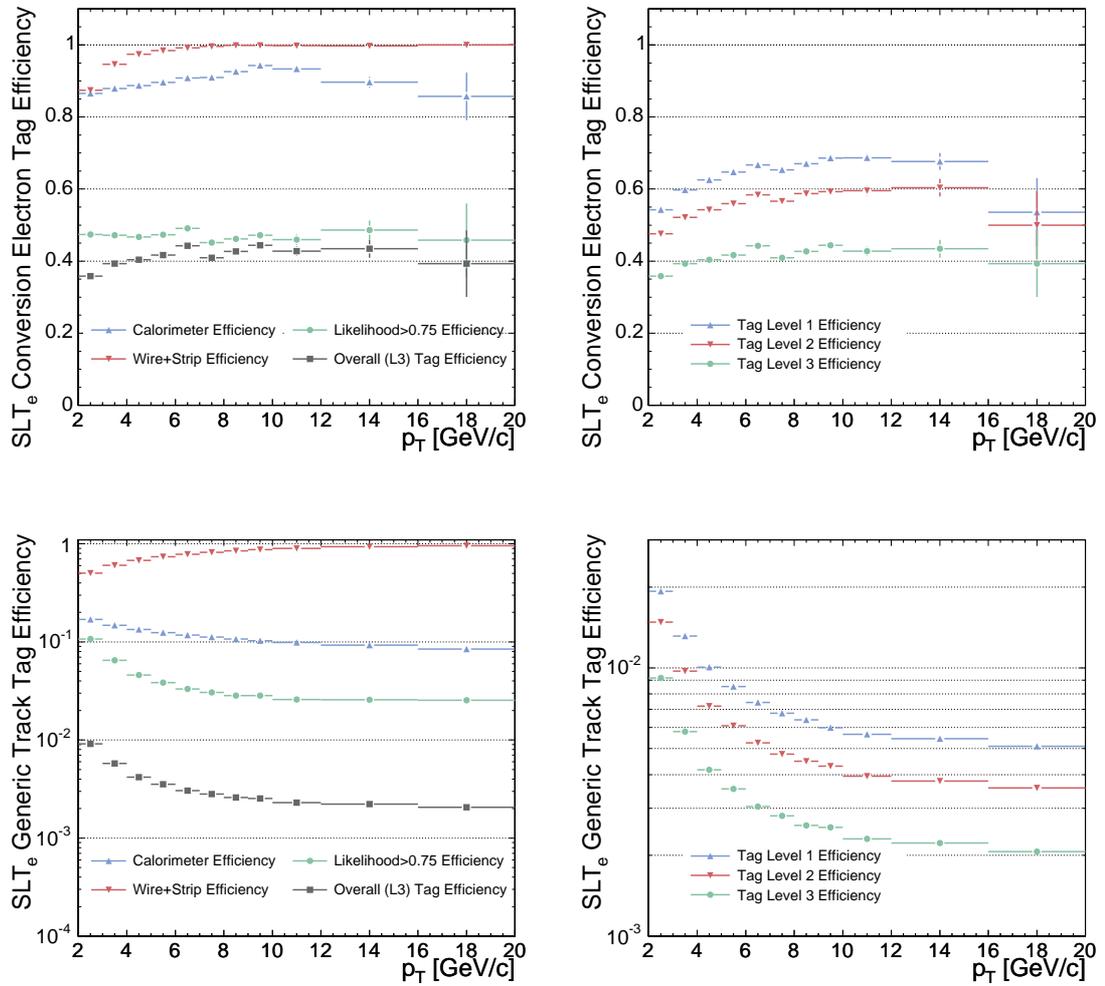


Figure 4.26: Tagging efficiency as a function of p_T for conversion electrons (top) and generic tracks (bottom). Shown is the efficiency of the individual requirements on the left, and for each tag level on the right.

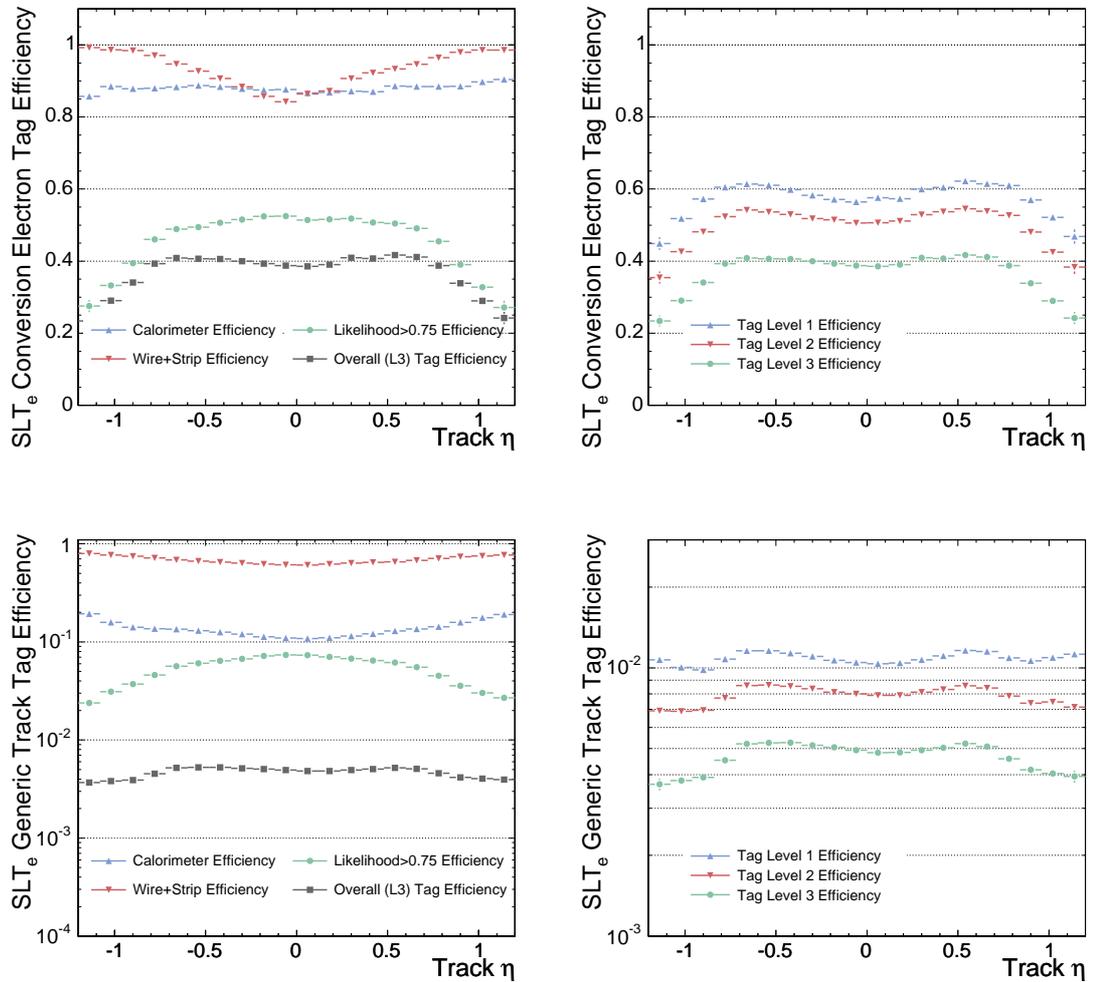


Figure 4.27: Tagging efficiency as a function of track η for conversion electrons (top) and generic tracks (bottom). Shown is the efficiency of the individual requirements on the left, and for each tag level on the right.

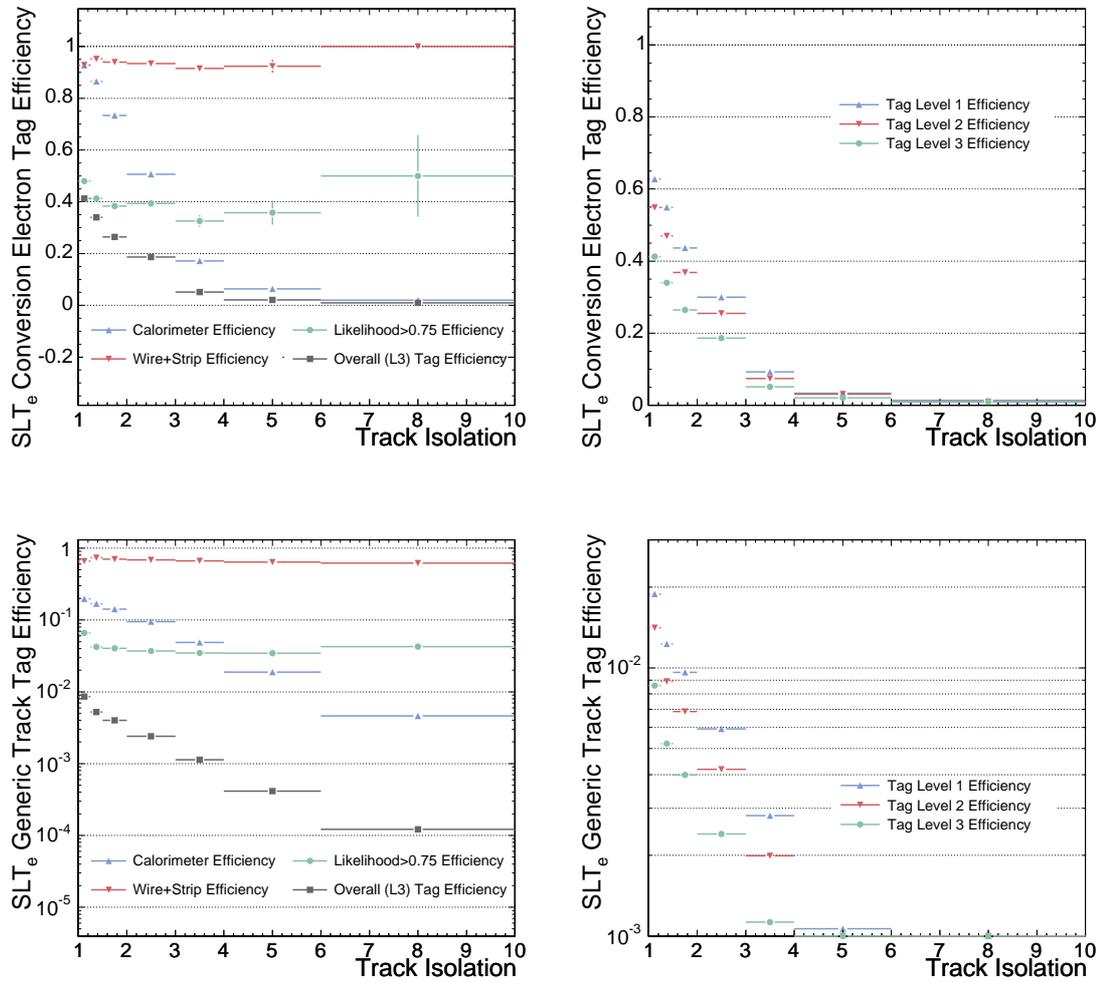


Figure 4.28: Tagging efficiency as a function of isolation for conversion electrons (top) and generic tracks (bottom). Shown is the efficiency of the individual requirements on the left, and for each tag level on the right.

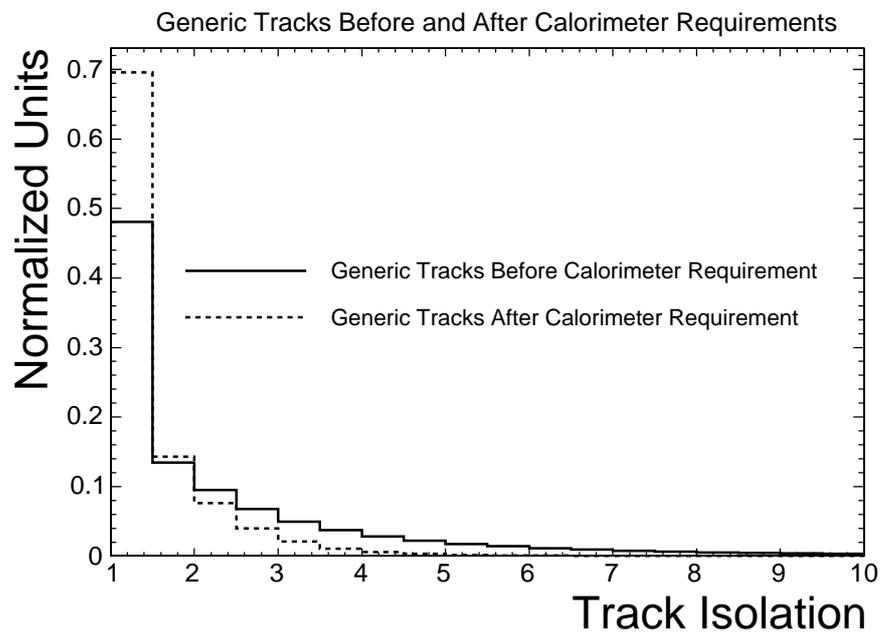


Figure 4.29: Comparison of the isolation distribution for the generic track sample before and after the E_{EM}/p and E_{Had}/E_{EM} requirements.

Chapter 5

Conversion Identification

5.1 Introduction

Although the SLT_e readily distinguishes between real electrons and fake electrons, it cannot classify electrons by their origin, in particular whether the electron comes from the semi-leptonic decay of HF or from a photon conversion. Some conversion rejection is possible with the tagger alone because the calorimeter requirement $0.6 < E_{EM}/p < 2.5$ can reject conversions when both legs deposit their energy in the same calorimeter towers. However, since conversion electrons are such a prominent background, this is problematic. For example, in $t\bar{t}$ MC events roughly 5 to 6 times as many central electrons embedded in jets with $p_T > 2 \text{ GeV}/c$ are from conversions as from HF decay. This ratio reduces to three conversion electrons per one HF electron by making a tight impact parameter requirement, $|d'_0| < 0.3 \text{ cm}$, but obviously a separate method for distinguishing between HF electrons and conversion electrons must be implemented.

Typically, a large fraction of conversions may be removed from an event by looking for another oppositely signed track near the candidate electron track. For example, high E_T electrons (see Section 3.2.1) are rejected when another track in the event has

- opposite charge
- $|\Delta \cot(\theta)| < 0.04$
- $|sep| < 0.2$ cm
- trident veto

$\Delta \cot(\theta)$ is the difference in $\cot(\theta)$ between the track pairs, and sep is the distance between the tracks when they are parallel in the $r - \phi$ plane. sep is signed so that it is negative if the tracks cross, and positive if they do not. The trident veto simply requires that there is not a third track in the event that also passes the criteria above. This can occur when a hard electron radiates a photon which promptly converts. We use this algorithm design as a starting point to produce our own conversion filter, but we improve upon it to accommodate the particularities of the soft electron environment (namely, low p_T and non-isolated). This improved filter has been discussed already in Chapter 4, but we explain its construction here.

The first improvement to the algorithm is to sign the conversion radius, R_{conv} . R_{conv} is calculated by measuring the distance from $(0, 0)$ to the point when the two tracks forming the conversion are parallel. For conversions, this variable is an accurate estimate of the actual conversion location of the photon. If R_{conv} originates behind $(0, 0)$ (analogous to SECVTX mistags), then we give the R_{conv} a negative sign and do

not consider the tracks as a conversion candidate. An illustration of this is shown in Figure 4.6.

Another improvement is to identify conversion candidates by whether or not an electron's track is expected to have, but does not have, silicon hits attached. Photons which convert well into the detector may produce electron tracks with a reduced number of silicon hits. By identifying such tracks, we can reject conversion electrons even when the partner leg is not reconstructed or lies outside the $\Delta \cot(\theta)$ and sep requirements. This is especially important at low p_T , because the track reconstruction efficiency declines at low p_T reducing the effectiveness of the filter. This effect is exacerbated by the asymmetric energy sharing between the electron-positron pair shown in Figure 5.1. Considering electrons with P_T as low 2 GeV/ c results in many conversion electrons without a reconstructed partner track.

We construct a variable, expected silicon layers, by extrapolating the track through the silicon detector and determining which ladders are intersected by the track. If a ladder is both intersected and properly read out, then that ladder is expecting a hit. Expected silicon layers is the number of layers where both axial and stereo sides of a silicon layer are expecting at least one hit. Due to the geometry of the silicon detector, some tracks can intersect the overlap of two wedges of the same layer. See, for instance, Figure 2.6. When no hits are attached to *either* side, then that layer is considered missing.

In summary, we reject tracks as candidates for SLT_e tagging if we find another oppositely-charged track in the event with $|\Delta \cot(\theta)| < 0.03$, $|sep| < 0.2$ cm, and $R_{conv} > 0$. The $\Delta \cot(\theta)$ requirement was tightened relative to the high E_T electron

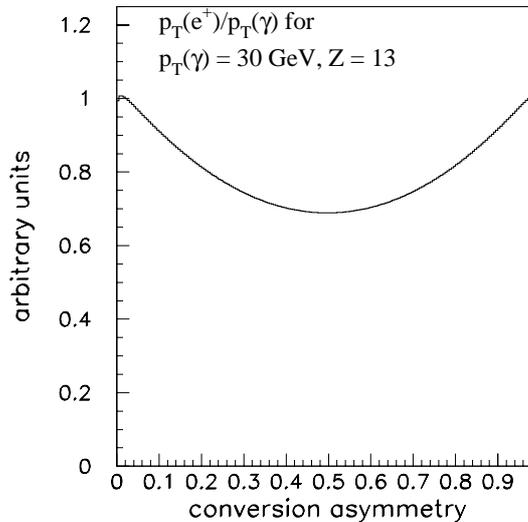


Figure 5.1: Analytic calculation of conversion asymmetry distribution for a photon with P_T of 30 GeV/c and material with atomic number, $Z=13$. The distribution represents the differential probability that the positron will receive a given fraction of the photon’s energy. Image taken from [1].

filter to reduce the rate which non-conversion tracks were misidentified as conversions. If no such track is found, we can still reject the track if it has more than three missing silicon layers. This filter is summarized in Table 5.1. We describe the missing silicon layer in greater detail in Section 5.3.

In Sections 5.4 and 5.5 we present a method of measuring the over-efficiency and efficiency of our filter, respectively. The over-efficiency is defined as the rate at which electron candidates are misidentified as having originated from a photon conversion. Our approach is data driven, relying on MC principally to estimate systematic uncertainties and verify the robustness of our method. The data driven aspect is important for conversions because of MC mis-modeling of the tracking reconstruction and the material description of the detector. Section 5.6 provides a cross check for our method

Conversion Filter Requirements
Reject tracks that have a partner track passing the requirements:
opposite charge
$ \Delta \cot(\theta) < 0.03$
$ sep < 0.2$ cm
Signed $R_{conv} > 0$ cm
Also reject tracks with > 3 Missing SI layers

Table 5.1: Final conversion filter applied to each candidate electron track in an event. Tracks are removed from consideration when either a partner track is found or too many silicon layers are missing.

in MC. By comparing the efficiency and over-efficiency in data and MC, we derive a Scale Factor in Section 5.7, which will be an important component of the SLT_e algorithm.

5.2 Sample Selection

In this section, we describe the construction of three samples: electron tags, anti-electron tags, and conversions. These samples are constructed in events from a variety of trigger paths. We use the inclusive ($E_T > 8$ GeV) electron trigger as well as the generic jet (jet20/50/70/100) triggers, which trigger on a jet with E_T greater than the corresponding amount. Each dataset from these triggers has its benefits to our study. The inclusive electron dataset has many more electrons than the generic jet datasets and so provides a more instructional starting place to study conversions. On the other hand, the generic jet datasets have conversions that are kinematically and environmentally similar to those in $t\bar{t}$ events, specifically, low p_T conversions embedded in jets. A second important feature of the generic jet datasets is the

availability of the appropriate corresponding MC which we use to measure a data-MC efficiency and over-efficiency SF.

5.2.1 Electron Tag Selection

Electrons are selected using the SLT_e tagger. The taggable selection requirements for electrons is identical to those in Table 4.6 except that the conversion requirements are not implemented. Since we want a pure sample, we require a level 3 SLT_e tag. We modify the selection for electrons in the inclusive electron dataset to preserve statistics and eschew a potential (albeit ultimately irrelevant) trigger bias. In this case, we require that the tracks have a p_T greater than 8 GeV/ c , and we no longer require that the tracks be embedded in a jet.

In the generic jet samples, we do not consider tracks matched to the trigger jet. The trigger jet is determined to be the closest jet in $\eta - \phi$ space to the Level 2 cluster above the trigger threshold. If more than one such jet exist, then we reconsider all of the tracks in the event.

5.2.2 Anti-Electron Tag Selection

Our approach here is to assume that the electron tag samples consist of either ‘prompt-like’ or ‘conversion-like’ elements. In general, conversions are distinguishable from other electrons because they often have large ‘decay’ radii due to material interactions and have small opening angles between partner tracks. Prompts, on the other hand, typically originate close to the origin and have large opening angles between themselves and other tracks in the event.

A few points complicate this simple model. The electron tag sample is not composed purely of electrons but contains some (sample dependent) amount of fake electrons as well. Therefore, ‘prompts’ can run the gamut from things such as pions from the hadronization of light jets to HF electrons. A pion that fakes an electron and is produced in the decay $K_s \rightarrow \pi^+\pi^-$ can look suspiciously like a conversion electron on two fronts: the long lifetime of the K_s results in decay vertices far removed from beam, and the high boost typical of K_s ’s produce a small opening angle between pions. Electrons from the Dalitz decay of neutral pions ($\pi^0 \rightarrow e^+e^-\gamma$) also look like conversions because of the small opening angle and can even have a large decay radius if the π^0 results from another long-lived particle.

In summary, we define prompts to be any track that is *not* a conversion electron. Prompts can have a wide range of behavior (from the perspective of the conversion identification) which must be accounted for. Of course, many fake conversions are backgrounds to the SLT_e tagger anyway, and so many be removed without consequence to the overall acceptance for HF electrons.

We choose to model prompts with anti-electron tags. Anti-electrons are those tracks which extrapolate to the central calorimeter and CES, but have some reversed electron identification requirements. These requirements are summarized in Table 4.8. As noted there, those requirements result in an extremely pure sample. We assume that HF electrons behave similarly to anti-electron tags with respect to the conversion filter. We will determine the validity of this assumption in Section 5.4.

5.2.3 Pure Conversion Electron Selection

Sample Construction

The conversion sample is a subset of the electron tag sample and is collected identically in events collected from the inclusive electron trigger and the generic jet triggers. We begin with the electron tag sample described above and look for another oppositely charged track in the event with $|\Delta \cot(\theta)| < 0.03$, $|sep| < 0.2$ cm, and $R_{conv} > 0$. We then tag the partner track with a level 3 SLT_e tag. Unlike the electron tag sample, the conversion sample is pure due to the fact that we tag both legs of the conversion. We will measure the purity shortly.

Table 5.2 summarizes the construction of electron tag, anti-electron tag, and conversion electron samples described in this section. These samples are constructed using events collected from the inclusive electron and generic jet triggers. Note that the conversion electron sample construction is different than what is implemented by the conversion filter, summarized in Table 5.1.

Sample Purity

As mentioned earlier, we give R_{conv} a negative sign if the conversion “originates” behind $(0,0)$. In Figure 5.2 we compare the radius of conversion distribution for both conversions and anti-electron tags in events collected from the inclusive electron trigger without applying the $R_{conv} > 0$ requirement. Since anti-electrons are pure fake electrons, they obviously do not have a true conversion radius. Shown then is the measured R_{conv} when a prompt is accidentally matched with an unrelated track. We note that, unlike in the conversion sample, there is a significant excess of anti-

Sample Selection Summary		
Electron Tag Sample	Anti-Electron Tag Sample	Conversion Sample
SLT _e taggable (Table 4.6)	SLT _e anti-tag (Table 4.8)	Electron Tag
No conversion requirement	No conversion requirement	Conversion Partner Found
Level 3 Tag (Table 4.7)		Partner has Level 3 Tag
Inclusive Electron Dataset Requirement Modifications:		
8 GeV/ <i>c</i> Track		
No Jet Matching Needed		
Generic Jet Dataset Requirement Modifications:		
Track not matched to trigger jet unless two jets pass trigger		

Table 5.2: Summary of the sample selection described in this section. These samples are taken from events in which the inclusive electron trigger or the generic jet triggers were fired.

electrons with negative R_{conv} . Although approximately 3% of the conversion sample has a negative R_{conv} , anti-electrons are nearly symmetric in this variable.

Whereas the probability of mis-tagging both legs and then pairing them as a conversion is quite low, other backgrounds do contribute. For example, electron/positron pairs from $\pi^0 \rightarrow e^+e^-\gamma$ may pass the conversion requirements, and populate the negative R_{conv} tail in approximately symmetric fashion about $R_{conv} = 0$. Additionally, semileptonic b decays where another random track fakes an electron and gets paired with the HF electron can also contribute to the conversion sample.

Fortunately, since prompts (both fake electrons and HF electrons) are approximately symmetric in R_{conv} about 0, we can use the size of the negative R_{conv} tail in conversions as a rough estimate of the total prompt contamination in the conversion sample with $R_{conv} > 0$ cm. This is only a rough estimate because real conversions can have a negative R_{conv} due to resolution effects.

Approximately 3% of the conversion sample has a negative R_{conv} , regardless of

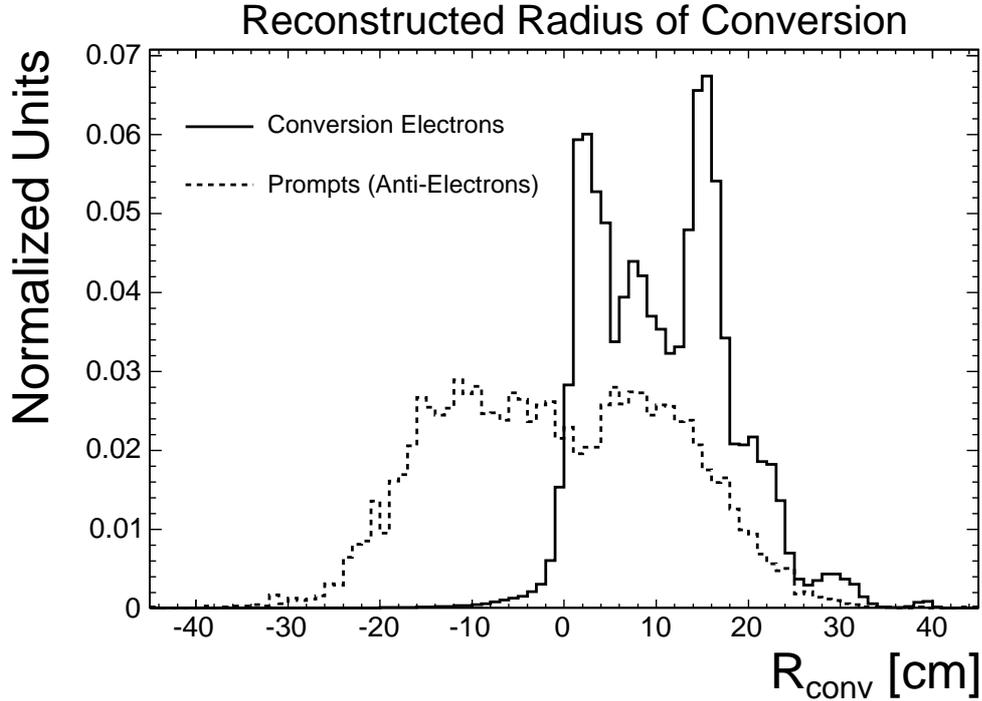


Figure 5.2: Radius of conversion for real conversions (top) and prompts modeled by anti-electrons (bottom). Conversion radius for anti-electrons can only be determined if another track is mistakenly paired with the original track. Conversions at high R_{conv} , such as at the COT inner cylinder, are suppressed due to the d'_0 cut.

the trigger used to select the event. We can examine the negative R_{conv} tail in MC, as well. In di-jet MC with at least one 50 GeV jet, the negative R_{conv} tail accounts for $(2.3 \pm 0.5)\%$ of the conversion sample, whereas, $(2.7 \pm 0.4)\%$ of conversion candidates with a positive R_{conv} are in fact due to non-conversions (mostly Dalitz decays). Therefore, we confirm that the negative R_{conv} distribution gives a proper estimate of the prompt contamination of the conversion sample. We quote the purity of the conversion sample in all datasets as $(97 \pm 1)\%$.

5.3 Missing Silicon Layers Description

Photons often convert well beyond the beam position and even beyond the silicon detector. We can exploit this feature by determining whether a track is expected to have a certain number of silicon hits attached. An excess of missing silicon hits will help identify conversion electrons even when their partner is not reconstructed.

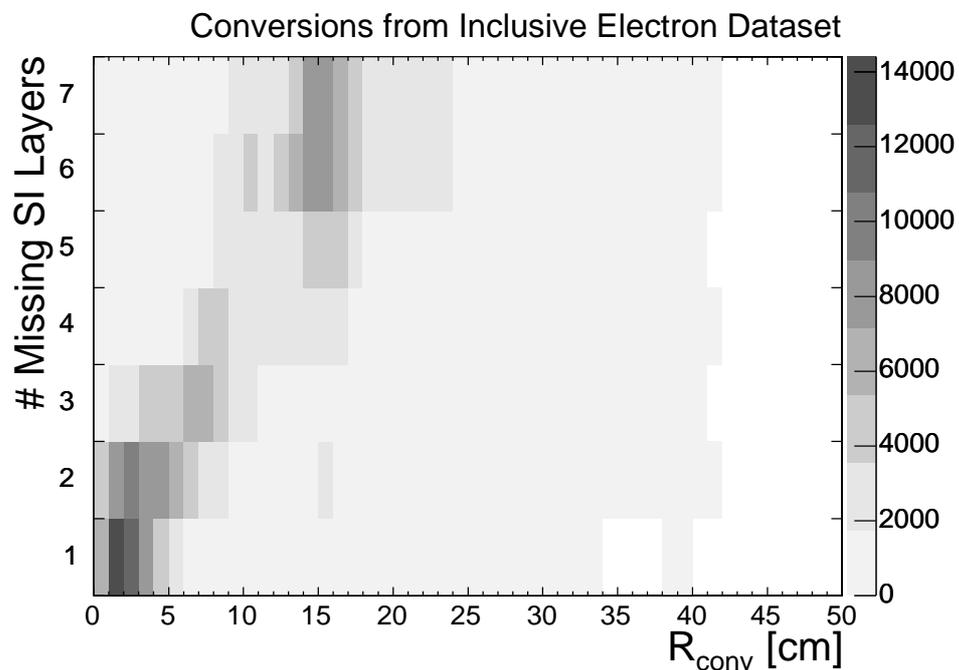


Figure 5.3: Missing silicon layers versus the radius of conversion for a sample of conversions. The SVX detector ends at approximately 12 cm. The suppression of conversions at high R_{conv} is due to the tight impact parameter requirement.

We plot the number of missing silicon layers for conversions from the inclusive electron dataset as a function of the radius of conversion in Figure 5.3. Here, we have required that $R_{conv} > 0$. It is clear that conversions display a strong correlation between missing silicon layers and R_{conv} . A couple points are also worth noting. First,

the number of expected/missing layers is at most six because L00 is single sided, and the second ISL layer is too forward in η to be fiducial for central electrons. That leaves the five layers of the SVX and the first ISL layer which can contribute to the missing layers variable. Second, the default track reconstruction algorithm heavily influences the missing silicon hits distribution. A minimum of 3 $r - \phi$ silicon hits must be attached to any track in order for it to pass the default tracking selection (See Section 3.1); otherwise, only COT hits are used. For example, a track can only be expecting six layers but missing five if two hits are found in L00¹. To show the full range of possibilities, Figure 5.4 shows the number of missing layers in different bins of expected layers for conversions and anti-electron tags from events triggered on an inclusive electron.

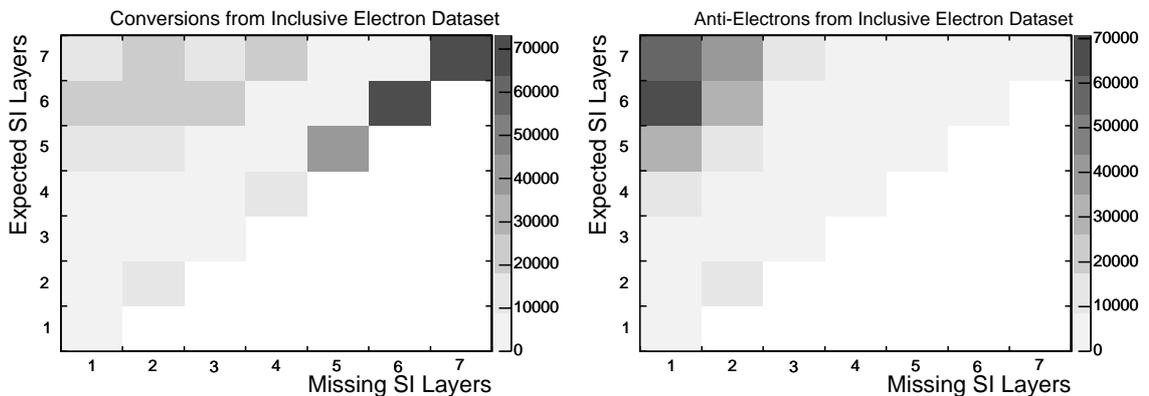


Figure 5.4: Missing silicon layers versus expected silicon layers for conversion electrons (left) and anti-electrons (right) in the inclusive electron dataset. The distribution of expected silicon layers is dependent on the geometry of the track.

To reduce the conversion background, we require that all candidate tracks have

¹This configuration is possible because L00 has significant geometric overlap. As a reference, see Figure 2.6.

≤ 3 missing silicon layers. In the following section, we will show that this results in approximately a 30% reduction of the tagging *inefficiency* in the 8 GeV electron dataset. At the same time, the over-efficiency nearly doubles.

5.4 Conversion Over-Efficiency

In this section we measure the over-efficiency of the conversion filter using events gathered from the inclusive electron trigger. The over-efficiency of filter in the generic jet samples will be measured later using the method explored here. The efficiency measurement depends on the over-efficiency, so we begin with the latter first.

5.4.1 Over-Efficiency Measurement

We measure the over-efficiency by applying the conversion filter to the anti-electron tag sample. The fraction of anti-electrons which fail the filter requirements is the over-efficiency. We make the assumption here that the over-efficiency for anti-electrons is the same as for generic prompts, an unspecified admixture of fake electrons and real electrons not from photon conversions.

Figure 5.5 shows the over-efficiencies as a function of p_T , η , z_0 , isolation, number of tracks, and the number of z vertices. We decompose the conversion filter into Partner Matching (PM) and Missing Silicon Layer (MSL) requirements to show the contributions from each. The average over-efficiency in the inclusive electron sample is (statistical errors only):

- partner matching over-efficiency: $(3.90 \pm 0.04)\%$

- missing silicon layer over-efficiency: $(4.07 \pm 0.04)\%$
- combined over-efficiency: $(7.81 \pm 0.10)\%$.

A few trends are worth noting. The over-efficiency for both MSL and PM requirements increase as a function of p_T . This feature is due to the correlation between track p_T and the boost of the decaying parent particle. A boosted particle will have a narrower opening angle between its daughter tracks and will traverse more of the detector before decaying.

Another feature is the variation of the MSL requirement efficiency as a function of η and z_0 , which is due to varying silicon detector coverage in those variables. The MSL requirement over-efficiency continues to rise as a function of the number of tracks in the event. Presumably, this is due to a loss of efficiency to recover silicon hits in a denser environment, although the correlation with the number of z vertices is weaker.

5.4.2 Over-Efficiency Systematics

Here we discuss our assumption that prompts can be modeled by anti-electrons. In particular, the signed radius of conversion and missing silicon layer distributions are key aspects of conversion identification. We can also appeal directly to MC (specifically dijets with a jet with $E_T > 50$ GeV) to determine whether anti-electrons and HF electrons have the same over-efficiency. Because the SLT_e tag requirements are not well-modeled in MC, instead we apply the taggable requirements (Table 4.6), and then determine the particle identification directly from the MC.

The average reconstructed R_{conv} is 19% higher for HF electrons than hadrons in the di-jet MC. This is due to the, on average, long lifetime of the HF jets. Similarly,

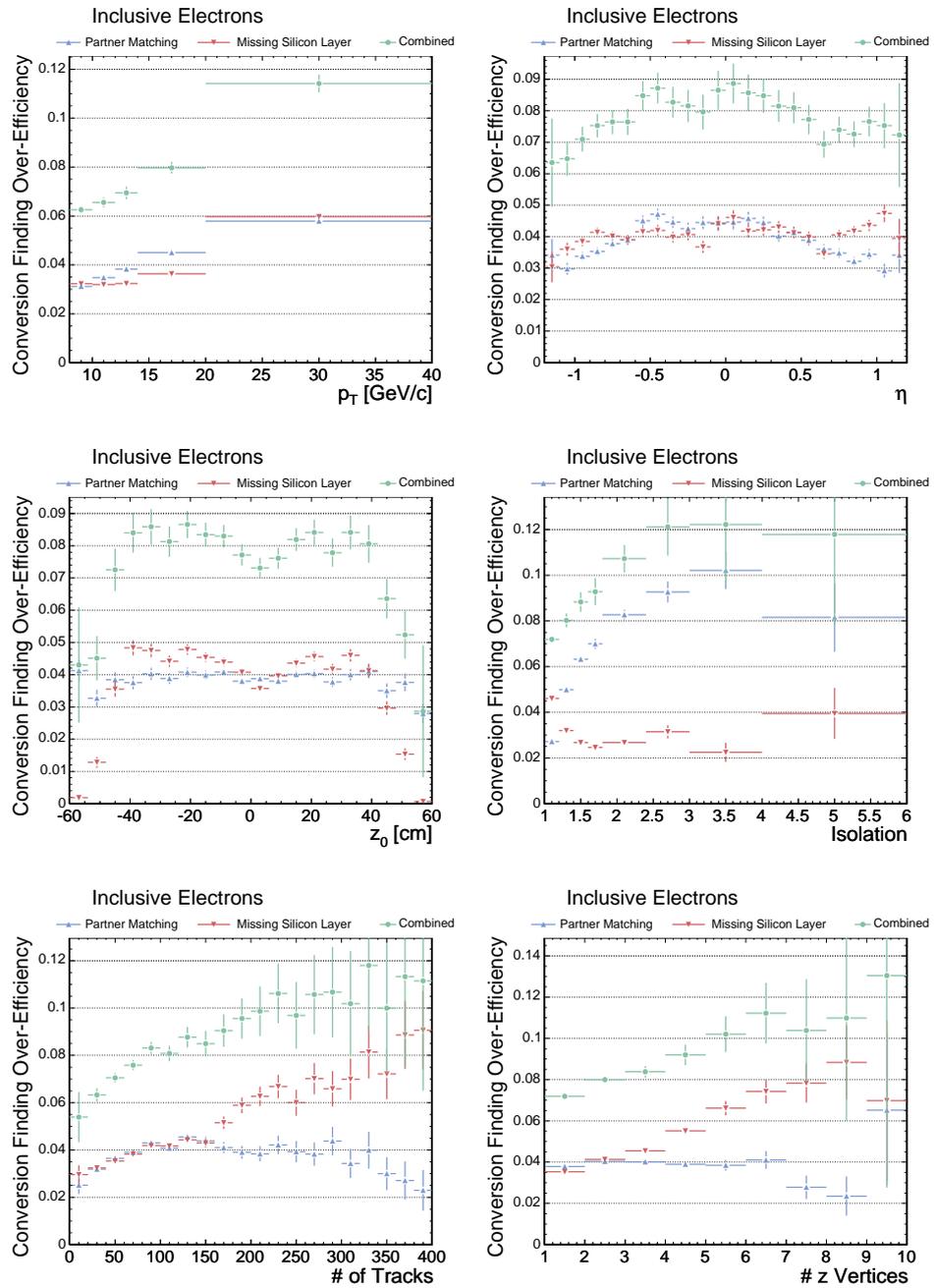


Figure 5.5: Over-efficiency as a function of the electron track p_T , η , z_0 , isolation, the number of tracks in the event, and the number of z vertices measured in events from the inclusive electron dataset.

the missing silicon layer requirement is approximately 61% more over-efficient for hadrons than for HF electrons. This is primarily due to the presence of particles with very long lifetimes, such as K_s and Λ .

Conveniently, the differences between hadrons and HF electrons with respect to the negative R_{conv} and missing silicon layer distributions contribute in opposite ways so that the overall difference in over-efficiency between the two is not as significant. The reason is that the negative R_{conv} requirement is sensitive to the mean decay length, whereas the missing silicon layer requirement is more sensitive to the tail.

The total over-efficiency of hadrons and HF electrons in di-jet MC is $(7.83 \pm 0.06)\%$ and $(5.69 \pm 0.35)\%$, respectively. The full 27% relative difference is applied as a systematic to the measured over-efficiency in the anti-electron sample. However, when either the track matching over-efficiency or missing silicon layer over-efficiency is used in isolation (as we will do in the following section), the proper corresponding systematic must be applied. For the inclusive electron dataset, we quote a combined over-efficiency: $\varepsilon^o = 7.8 \pm 0.1$ (stat) ± 2.1 (syst) (%).

5.5 Conversion Efficiency

As in the previous section, we measure the conversion efficiency in events collected from the the inclusive electron trigger. We measure the efficiency of the conversion filter by decomposing the filter into Partner Matching and Missing Silicon Layer requirements and treating each component separately. We recombine them adjusting for their correlations.

5.5.1 Partner Matching Efficiency Measurement

We can measure the PM requirement efficiency if we know how many conversion electrons and prompts make up the electron sample before the PM requirement is applied. The number of electrons tags that we identify as conversions through the PM requirement is equal to the original number of conversion electrons in the sample (before conversion removal) times the efficiency plus the number of prompts (again before conversion removal) times the over-efficiency:

$$N_{ele}^{id} = N_c \varepsilon + N_p \varepsilon^o. \quad (5.1)$$

Fortunately, we have a convenient handle to measure the fraction of conversions and prompts that make up the electron sample: the missing silicon layer variable. By constructing missing silicon layer templates out of the conversion and anti-electron samples, we can fit them to the electron tag distribution to determine the contribution from each. In the fit we use the templates when six silicon layers are expected. The fitted fraction of electrons that are from conversions in the inclusive electron dataset is $(53.0 \pm 0.2)\%$, shown in Figure 5.6.

The fit is not perfect because the conversion MSL distribution is sculpted by the reconstruction and tagging of both legs of the conversion. We address this issue in the next section, but do not correct for it here. The fitted value also needs to be corrected for prompt contamination of the conversion sample ($(3 \pm 1)\%$ as determined earlier in Section 5.2). The corrected equations are given by

$$N'_p = (N_c + N_p)(1 - F_{conv}(1 - R)) \quad (5.2)$$

$$N'_c = (N_c + N_p)F_{conv}(1 - R) \quad (5.3)$$

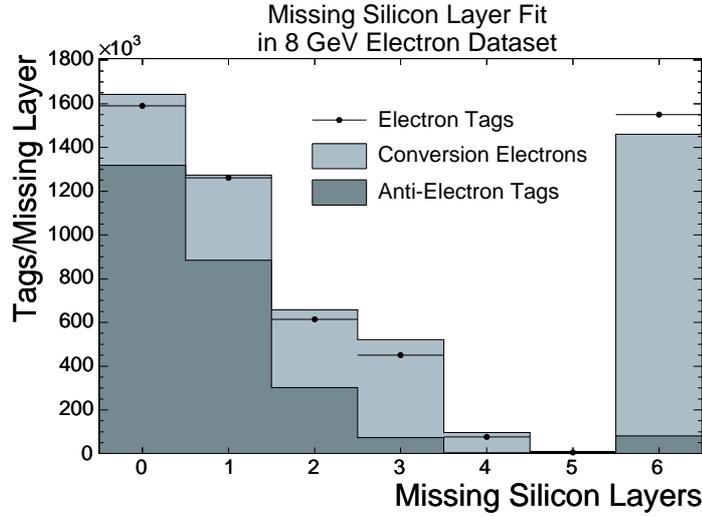


Figure 5.6: Fit result for missing silicon layer templates.

$$\varepsilon' = (N_{ele}^{id} - N_p' \varepsilon^o) / N_c' \quad (5.4)$$

where the primed variables are those corrected for the conversion sample contamination by prompts. F_{conv} is the fitted conversion fraction, and R is the contamination estimate ($\sim 3\%$). Note that $N_c + N_p$ is the total number of electron tags before conversion removal. With $N_c + N_p = 15.9 \times 10^6$, and $N_{ele}^{id} = 5.9 \times 10^6$, we get

$$N_p' = 7.73 \times 10^6 \quad (5.5)$$

$$N_c' = 8.16 \times 10^6 \quad (5.6)$$

$$\varepsilon' = (68.8 \pm 0.3)\% \quad (5.7)$$

where we are giving the statistical and fit errors only.

We note in summary that three systematic uncertainties need to be accounted for in this part of the measurement. First is the systematic uncertainty on the conversion sample purity. The second is the systematic uncertainty on the partner matching

Expected SI Layers	Measured Efficiency (%)
3	66.4 ± 0.1
4	66.9 ± 0.1
5	68.0 ± 0.2
6	68.8 ± 0.3

Table 5.3: Conversion finding efficiency by different number of expected silicon layers.

requirement over-efficiency used in Equation 5.1. We use the uncertainty derived in the previous section.

The third is based on the choice of using templates based on six expected silicon layers. Table 5.3 shows the conversion efficiency as a function of expected silicon layers. The efficiencies are consistent across different value of expected layers. The trend we observe is due to the fact that having six expected silicon layers implies a specific track geometry. We apply a three percent systematic to the PM requirement efficiency measurement. We quote for the inclusive electron sample: $\varepsilon_{PM} = 68.8 \pm 0.3$ (stat) ± 2.2 (syst)(%).

5.5.2 Missing Silicon Layer Efficiency Measurement

We now turn to the missing silicon layer requirement. To measure this efficiency, we simply apply the MSL requirement to the conversion sample. 44.9% of the conversions have more than three missing silicon layers. We again correct for the prompt contamination of the conversion sample: $\varepsilon' = \varepsilon/(1 - R)$, where again R is the contamination estimate. The corrected efficiency is $\varepsilon_{MSL|PM} = 46.2\%$. In principle, there is a second order correction due to the fact that prompts have a small over-efficiency.

This correction is on order of $\sim \varepsilon^o \cdot R$, which is sufficiently small that we can ignore it.

We have labeled the efficiency $MSL|PM$ since this is the efficiency to find too many missing silicon layers when a partner has already been found. In di-jet MC, the difference between the MSL efficiency with and without the partner leg present is appreciable. Approximately 47% of conversions with the partner leg found pass the MSL requirement in MC, whereas only 42% of *all* conversions pass the MSL requirement. Since in data we only measure the MSL efficiency when the partner leg is present, we reduce the MSL efficiency we measure by $(47 - 42/47) = (11 \pm 5)\%$ relative. This correction admittedly does not quite conform to our “MC is used for systematics only” approach. Nonetheless, the effect is large, and so we have introduced a correspondingly large systematic. We also check that the other generic jet MC datasets have a similarly large correction. Once we have accounted for this correction, we quote for the inclusive electron dataset: $\varepsilon_{MSL} = 41.2 \pm 0.1$ (stat) ± 2.1 (syst)(%).

5.5.3 Combined Efficiency Measurement

We combine the efficiencies taking into account the correlations. The total efficiency is

$$\varepsilon = \varepsilon_{MSL} + \varepsilon_{PM} * (1 - \varepsilon_{MSL|PM}). \quad (5.8)$$

Therefore, in the inclusive electron dataset, we measure a combined efficiency of $\varepsilon = 78.1 \pm 0.1$ (stat) ± 1.3 (syst)(%).

Figure 5.7 shows the conversion filter efficiencies as a function of p_T , η , z_0 , isolation,

the number of tracks in the event, and the number of z vertices. Each bin has been fitted independently with conversion and anti-electron templates taken from that bin only. Only the variation in the size of the negative R_{conv} tail as a function of each variable has not been taken into account. Although the uncertainties shown are technically uncorrelated, the fit uncertainty can be made larger if the missing silicon layer shapes are systematically different from the true shapes. The increasing PM efficiency with p_T is due to the increasing reconstruction efficiency of the conversion partner track. The MSL efficiency dependence on p_T seems to stem from a correlation with the radius of conversion.

As we have seen before with the over-efficiency, the variation of the MSL efficiency in η and z_0 is due to changing silicon detector coverage. Note also that the 3 barrel structure is clearly apparent in the MSL efficiency z_0 distribution. The missing efficiency at extreme values of z_0 is because of a geometric bias introduced by our use of missing silicon hit templates when all six layers are required to be expected. The decreasing efficiency as a function of the number of tracks and z vertices is unfortunately anti-correlated with the trend for the over-efficiency. In the case of the MSL requirement, it seems that in a higher density region, silicon hits get improperly assigned to the wrong tracks more often.

Two aspects of our approach bear discussion. First, our method is reliant on data because of differences in MC with track reconstruction and the detector material description. Second, the measurement uses templates built within the dataset itself rather than measuring templates in one sample and applying them in another. This *in situ* method is important because of the sample dependence of the conversion filter

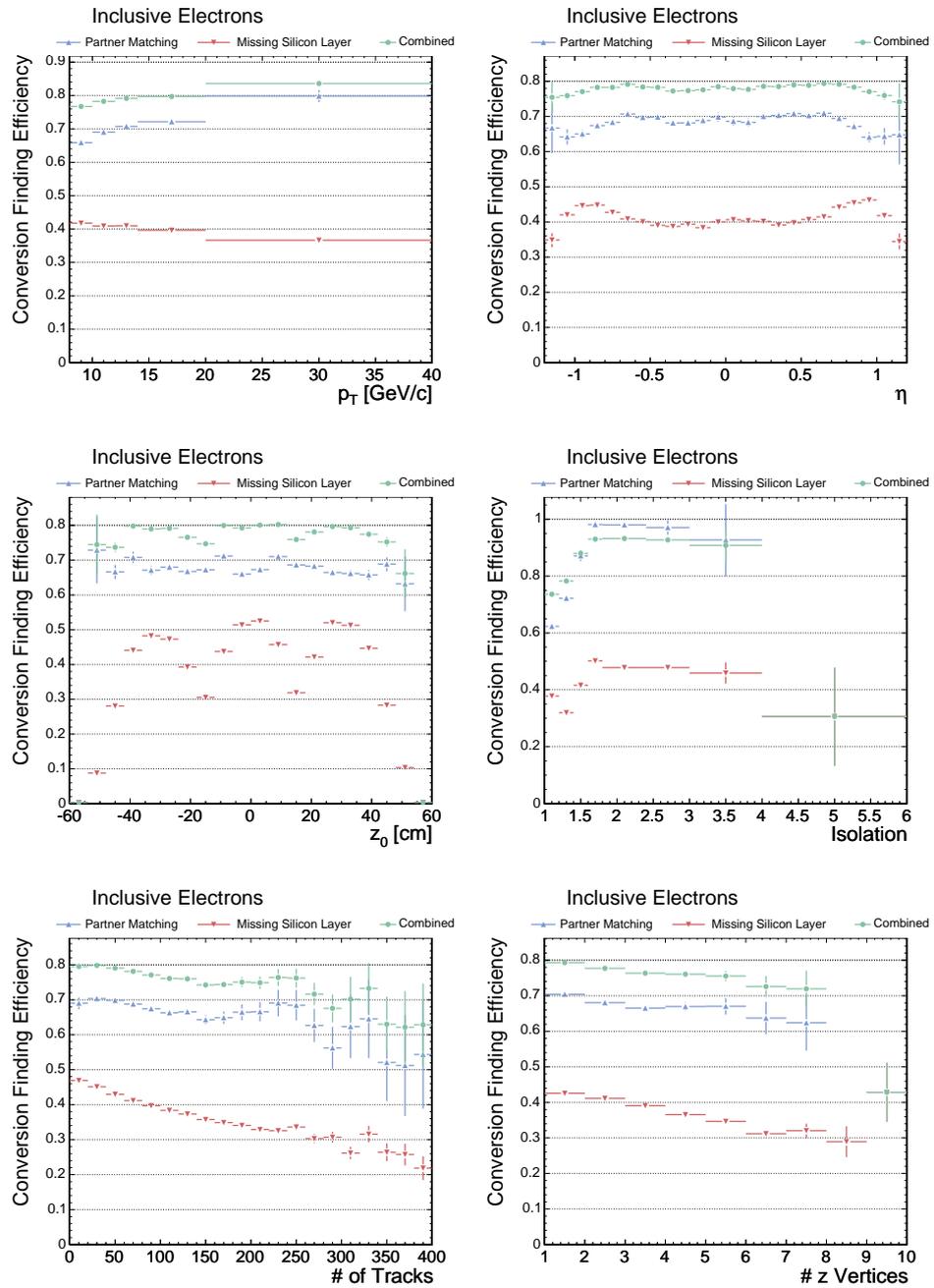


Figure 5.7: Efficiency as a function of the electron track p_T , η , z_0 , isolation, the number of tracks in the event, and the number of z vertices measured in events from the inclusive electron trigger.

efficiency. The efficiency is a strong function of the photon energy spectrum, which can vary greatly between different samples. Softer photons will yield softer conversion electron tracks, reducing the total conversion filter efficiency. This does not mean that the conversion efficiency can be applied across samples simply by parameterizing the efficiency by the candidate electron track p_T . The critical factor is the p_T of photon which is unknown unless the partner leg is identified. Another effect that is taken into account with this method is the geometric dependence of the silicon detector coverage. A bias could be introduced if the sample in which the conversion efficiency is measured is not, on average, the same in those respects as the sample in which it is applied.

5.6 Method Cross Check

As a final cross check to determine whether or not our method is reliable, we can apply the efficiency calculation in MC and compare against the true efficiency determined from MC. Table 5.4 shows the efficiency measured in the generic jet MC datasets using the same method that we have used to measure the efficiency in the inclusive electron sample. Again, we do not apply the full SLT_e tagger, but only the taggable requirements and then ‘tag’ using the true information from MC. The agreement is very good, within systematic uncertainties. Therefore, we are confident that the method we have undertaken properly estimates the efficiency in data.

Dataset	Measured Efficiency (%)	True Efficiency (%)	Δ (%)
jet20 MC	73.0 ± 1.8 stat ± 1.3 syst	76.5 ± 0.3	-3.5 ± 2.3
jet50 MC	76.0 ± 1.3 stat ± 1.3 syst	75.2 ± 0.2	0.8 ± 1.9
jet70 MC	72.6 ± 2.5 stat ± 1.3 syst	73.8 ± 0.4	-1.2 ± 2.8
jet100 MC	75.2 ± 2.9 stat ± 1.3 syst	73.6 ± 0.3	1.6 ± 3.2

Table 5.4: Cross check measuring the efficiency in the generic jet datasets and comparing them against the true efficiency.

5.7 Scale Factors

5.7.1 Efficiency Scale Factor

We measure the efficiency SF by measuring the efficiency in each of the generic jet datasets and measuring the true efficiency in the corresponding generic jet MC.

Tables 5.5, 5.6, and 5.7 show the breakdown of the SF measurement for each dataset as well as for the PM, MSL, and combined efficiencies. The SF is in excellent agreement between all four of the generic jet datasets. Therefore, we quote a total conversion efficiency data-MC SF:

$$SF = 0.93 \pm 0.01 \text{ stat} \pm 0.02 \text{ syst}. \quad (5.9)$$

It is interesting to see that the partner matching efficiency is only a little higher than the corresponding efficiency in MC. This suggests that the tracking differences between data and MC are not as great as we initially may have guessed. On the other hand, the missing silicon layer efficiency is very different between data and MC. This is not unsurprising since the material description in MC has known discrepancies with data. When combined together, the differences cancel somewhat to give a scale factor close to unity.

Partner Matching Efficiency Scale Factor			
Dataset	Efficiency in Data (%)	True Eff. in MC (%)	Data-MC SF
jet20	$60.9 \pm 2.1 \pm 1.9$	$55.5 \pm 0.4 \pm 0.0$	$1.09 \pm 0.04 \pm 0.03$
jet50	$61.9 \pm 2.5 \pm 1.9$	$57.2 \pm 0.2 \pm 0.0$	$1.08 \pm 0.04 \pm 0.03$
jet70	$59.5 \pm 2.2 \pm 1.9$	$57.4 \pm 0.4 \pm 0.0$	$1.03 \pm 0.04 \pm 0.03$
jet100	$60.7 \pm 2.5 \pm 1.9$	$57.8 \pm 0.4 \pm 0.0$	$1.05 \pm 0.04 \pm 0.03$

Table 5.5: Measurement of the Partner Matching efficiency in data and comparison against the true efficiency in the corresponding MC. The first uncertainty is statistical, the second is systematic.

Missing Silicon Layer Efficiency Scale Factor			
Dataset	Efficiency in Data (%)	True Eff. in MC (%)	Data-MC SF
jet20	$31.3 \pm 0.8 \pm 1.6$	$47.4 \pm 0.4 \pm 0.0$	$0.66 \pm 0.02 \pm 0.03$
jet50	$26.5 \pm 0.9 \pm 1.4$	$41.7 \pm 0.2 \pm 0.0$	$0.64 \pm 0.02 \pm 0.03$
jet70	$27.1 \pm 0.9 \pm 1.4$	$38.4 \pm 0.4 \pm 0.0$	$0.71 \pm 0.02 \pm 0.04$
jet100	$23.2 \pm 0.8 \pm 1.2$	$36.2 \pm 0.4 \pm 0.0$	$0.64 \pm 0.02 \pm 0.03$

Table 5.6: Measurement of the Missing Silicon Layer efficiency in data and comparison against the true efficiency in the corresponding MC.

Combined (PM + MSL) Efficiency Scale Factor			
Dataset	Efficiency in Data (%)	True Eff. in MC (%)	Data-MC SF
jet20	$70.7 \pm 1.4 \pm 1.3$	$76.5 \pm 0.3 \pm 0.0$	$0.92 \pm 0.02 \pm 0.02$
jet50	$70.0 \pm 1.9 \pm 1.4$	$75.2 \pm 0.2 \pm 0.0$	$0.93 \pm 0.03 \pm 0.02$
jet70	$68.5 \pm 1.6 \pm 1.4$	$73.8 \pm 0.3 \pm 0.0$	$0.93 \pm 0.02 \pm 0.02$
jet100	$68.1 \pm 1.6 \pm 1.5$	$73.6 \pm 0.3 \pm 0.0$	$0.92 \pm 0.02 \pm 0.02$
comb.	—	—	$0.93 \pm 0.01 \pm 0.02$

Table 5.7: Measurement of the combined conversion filter efficiency in data and comparison against the true efficiency in the corresponding MC.

Figures 5.8, 5.9, and 5.10 show the data-MC efficiency SF as a function of p_T , z_0 , and the number of tracks in the event for the generic jet datasets, jet20/50/70/100.

Note that in certain bins, the calculation does not have enough statistics to fit for the PM efficiency in data.

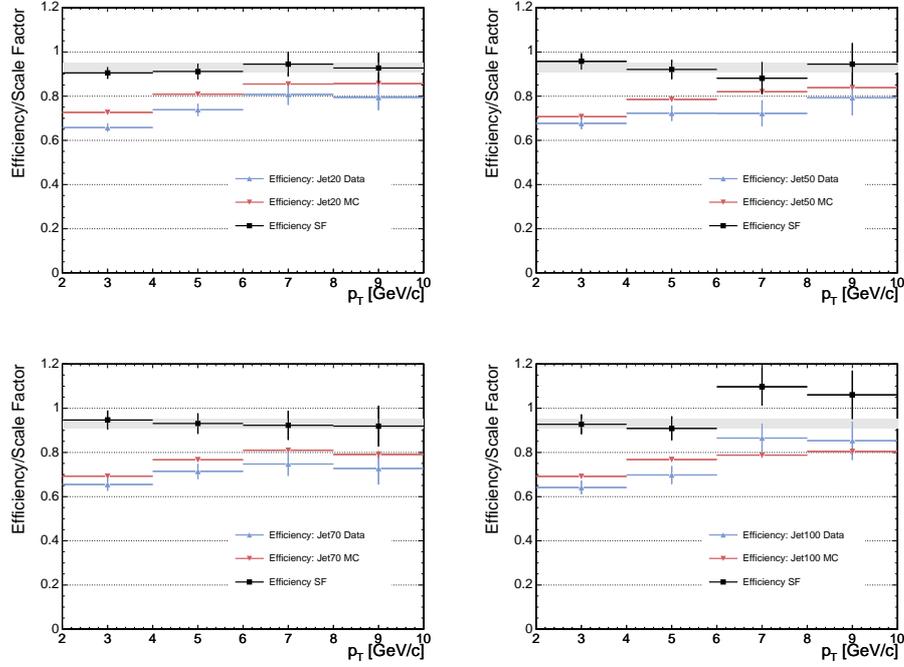


Figure 5.8: Data-MC Efficiency Scale Factor in the generic jet datasets as a function of p_T .

From these figures, we can verify that the SF is constant across different variables, encompassing kinematics, geometry, and environment. This indicates that the scale factor concept is robust. This is particularly important when considering the p_T dependence. If the photon spectrum in the MC were different than in data, it would be made manifest as a slope in the SF as a function of p_T . Since no discernable slope is shown, we are confident that the MC does predict the photon spectrum correctly.

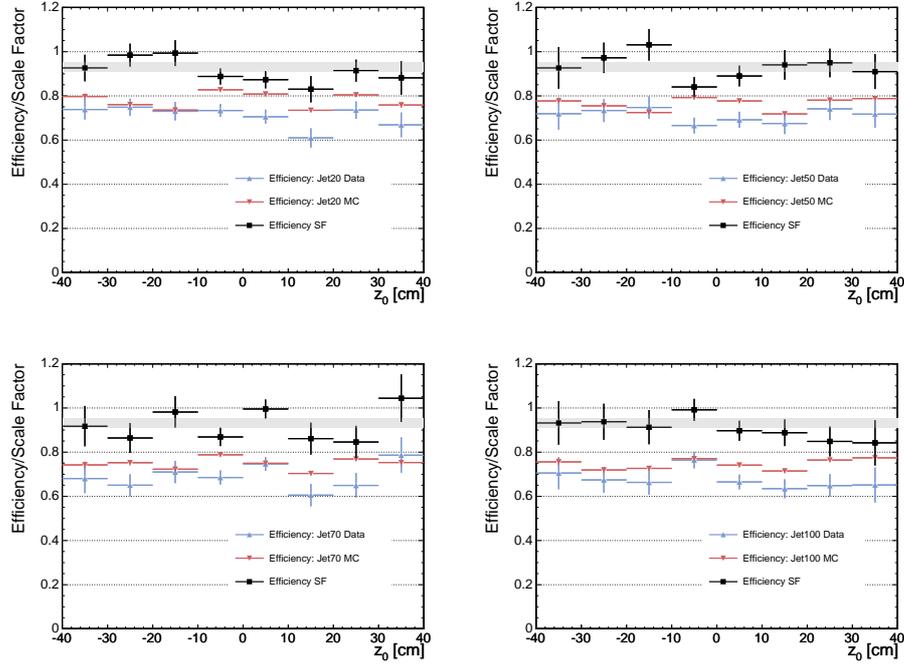


Figure 5.9: Data-MC Efficiency Scale Factor in the generic jet datasets as a function of z_0 .

5.7.2 Over-Efficiency Scale Factor

To measure the over-efficiency SF, we must first take into account the fact that the p_T spectrum of the fake sample in data is not the same as the p_T spectrum of the fake sample in MC. This is because the anti-tag requirements in data favor higher p_T tracks. Although the efficiency measurement is insensitive to this effect, it merits mentioning here.

Figure 5.11 shows the over-efficiency as a function of p_T in the different generic jet datasets. The measurement is a direct comparison between anti-electron tags in data and MC. We place a large systematic to cover the variation across p_T in all generic

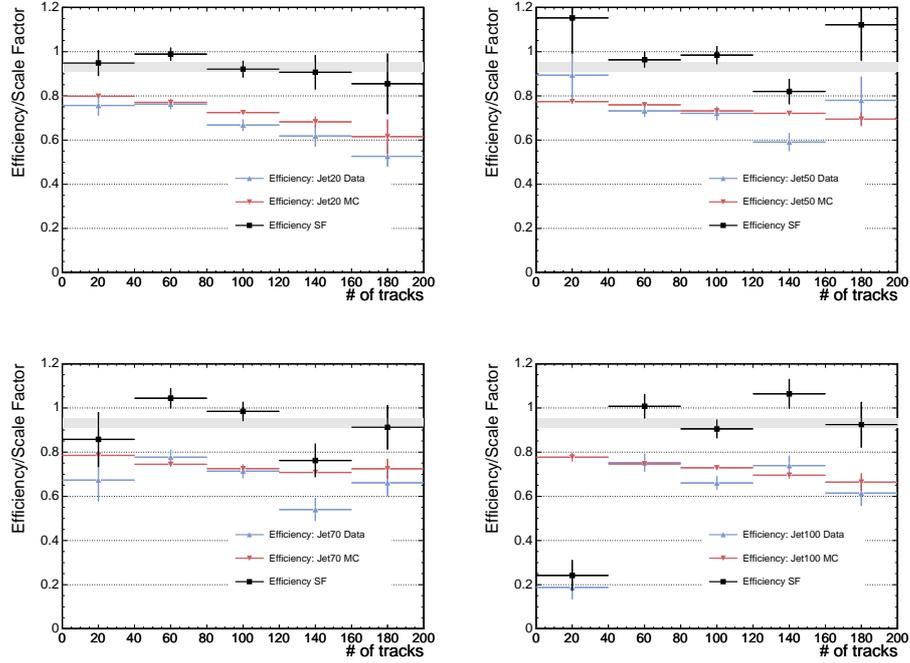


Figure 5.10: Data-MC Efficiency Scale Factor in the generic jet datasets as a function of the number of tracks in the event.

jet datasets. We quote the data-MC over-efficiency SF as:

$$SF = 1.0 \pm 0.3 \text{ syst.} \quad (5.10)$$

This is likely an over-estimate of the SF due to the p_T bias of the anti-electron tags.

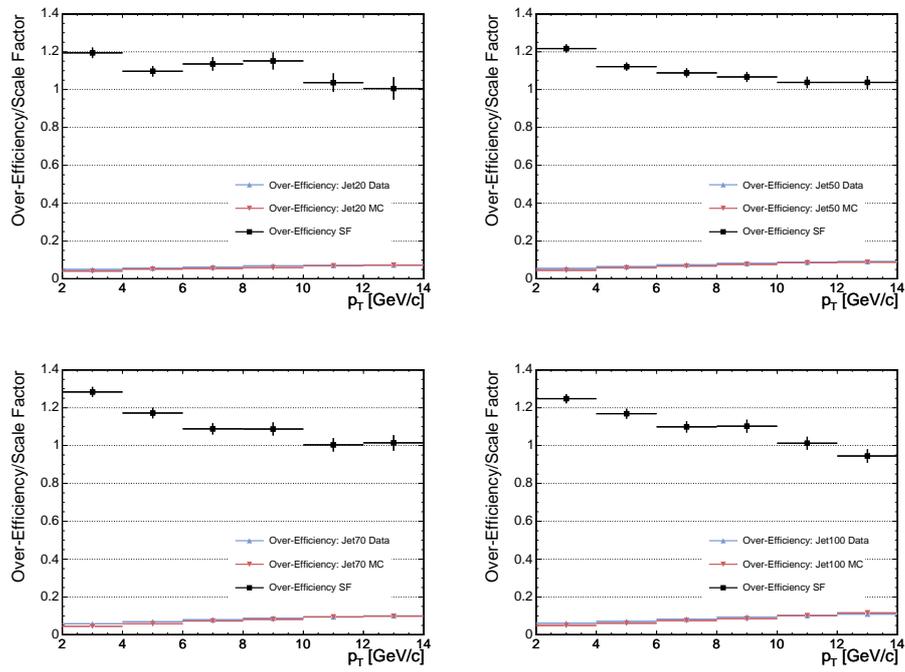


Figure 5.11: Data-MC Over-Efficiency Scale Factor in the generic jet datasets as a function of p_T .

Chapter 6

Tag Matrix

In this chapter, we return to our discussion of the SLT_e algorithm. Tables 4.6 and 4.7 summarize the taggable and tag SLT_e requirements, respectively. In this chapter, we find that the calorimeter requirements in Table 4.6, $0.6 < E_{EM}/p < 2.5$ and $E_{Had}/E_{EM} < 0.2$, seem to be well-modeled in MC, whereas the CES requirements are not. This fact motivates the definition of taggable to include the calorimeter requirements and the definition of tag to include only the CES requirements. Therefore, rather than apply the tagging requirements to the electron distributions in MC directly, we measure the efficiency - defined as *tagged over taggable* - of electrons in data, and apply that efficiency as a tagging probability to taggable tracks in MC. We call the efficiency parametrization as a function of p_T , η , and isolation the *tag matrix*. A separate matrix is constructed for each of the three tag levels of the SLT_e algorithm.

The parametrization variables are chosen because they characterize electron tagging by the kinematics, geometry, and isolation. We demonstrate in this chapter that

the tagging rate for soft electrons can be succinctly described as a function of these three variables within a few percent. This is done by comparing the measured and predicted tagging efficiency in different samples and across different variables.

We use the same pure conversion electron sample constructed in Chapter 4 to measure the tagging efficiency and build the tag matrix. Understanding the dependence on p_T and isolation is particularly important because we extrapolate the properties of low p_T , relatively isolated conversion electrons to high p_T , non-isolated heavy flavor (HF) electrons embedded in jets. We use a sample of electrons from Z bosons to cross check the tag matrix in a high p_T sample. However, we will wait until Chapter 8 to cross check the isolation extrapolation with a sample of b -jets when we have the appropriate machinery.

6.1 Tag Matrix Definition

We use electrons from the pure conversion sample described in Section 4.2.1. As described in that section, we estimate that this sample is better than 98% pure, so no background estimation is necessary. We measure the tagging efficiency

$$\varepsilon \equiv \frac{N_{tagged}}{N_{taggable}} \quad (6.1)$$

in different bins of p_T , η_{trk} , and isolation. The binning we choose for the tag matrix is:

- p_T : [2.0, 2.5, 3.0, 4.0, 5.0, 6.0, ∞)
- $|\eta|$: [0.0, 0.4, 0.8, 1.2)

- $I = 1, I \neq 1$

The primary concerns for the binning are minimizing statistical uncertainties while covering a wide range of phase space as finely as possible.

It turns out that the tag matrix can be factorized into a $p_T : \eta$ component and a $\eta : I$ component without losing its predictive ability. This is because the difference in tagging probability between the isolated and non-isolated case is roughly a function only of η . Therefore, we are able to gain in statistical power, at the low cost of having a non-trivial statistical error calculation.

In Figure 4.27, we see that the tagging efficiency has a large slope in the forward regions of η . Since we use coarse binning in η , the tag matrix will not predict the efficiency properly. Therefore, we correct the forward regions ($|\eta| > 0.8$) with a linear, multiplicative correction

$$1 - a \cdot |\eta| + b. \quad (6.2)$$

We leave the central regions uncorrected. The values for each tag level are given in the table below.

Tag Level	constant ‘a’	constant ‘b’
1	0.908	0.837
2	1.096	1.004
3	1.524	1.399

Table 6.1: Constants for the η corrections in the forward, $|\eta| > 0.8$, regions.

6.2 High p_T Extrapolation

Our first cross check of the tag matrix is in a high p_T sample of electrons from Z bosons. We trigger on a high E_T electron (see Section 3.2.1 for details on high E_T electron definition used) and select Z bosons by identifying an oppositely-charged SLT_e candidate in the proper invariant mass region. Just as in the case of the conversion electron sample described in Section 4.2.1, the partner leg is unbiased by the trigger. Partner legs must be taggable SLT_e tracks with a p_T greater than 20 GeV/ c , but no jet matching is required. The invariant mass between the trigger electron and the SLT_e electron is shown in Figure 6.1. Because the taggable definition includes calorimeter depositions consistent with an electron, we expect a pure sample of electrons from Z once we have selected the appropriate invariant mass window. This sample selection is summarized in Table 6.2.

Z Electron Sample Selection
Triggered High E_T Electron (See Section 3.2.1 for details)
SLT_e Taggable Tracks (See Table 4.6 for definition of taggable)
$p_T > 20$ GeV/ c
no jet matching
trigger electron and SLT_e candidate have opposite charge

Table 6.2: Z electron sample selection summary.

Since the invariant mass was calculated using the reconstructed track, a significant bremsstrahlung tail can be seen. At high p_T , bremsstrahlung becomes an increasingly important issue. High E_T photons can distort the CES variables. The tag matrix over-predicts the tagging rate substantially, particularly with the tighter SLT_e tag

levels which are more sensitive to slight variations in the distributions. Because the total bremsstrahlung component is un-known *a priori*, we consider both a tight ($80 < M_Z < 100 \text{ GeV}/c^2$) and a loose ($50 < M_Z < 100 \text{ GeV}/c^2$) requirement on the invariant mass window. In each window we measure the expected and predicted tag rates. We take the average relative difference between the measured and predicted tags in each mass window. We use average relative difference as a correction, and half difference of the corrections as a systematic uncertainty:

$$\text{Correction} = \frac{\Delta_l + \Delta_t}{2} \pm \frac{\Delta_l - \Delta_t}{2}, \quad (6.3)$$

where Δ_t and Δ_l are the relative differences between the predicted and measured tags using the tight and loose Z mass windows, respectively. Table 6.3 shows the results. We extend this correction down to $p_T > 16 \text{ GeV}/c$ where we see the beginnings of this distortion in the conversion sample.

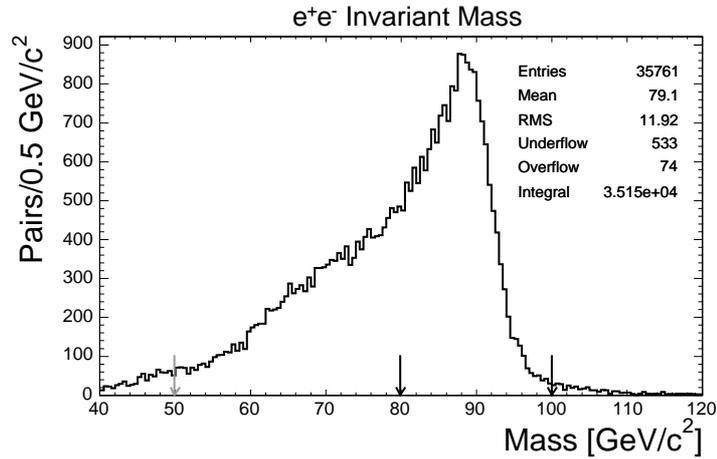


Figure 6.1: Invariant mass of high p_T trigger electron and candidate SLT_e track. The candidate track must have opposite sign, $p_T > 20 \text{ GeV}/c$, and pass taggable requirements.

80 < M_Z < 100 (GeV/ c^2)				
Tag Level	Measured	Predicted	$\Delta_t = (P - M)/P$	
1	13677 \pm 116.9	14100.5 \pm 76.3	(3.0 \pm 1.0)%	
2	11733 \pm 108.3	12296.4 \pm 82.9	(4.6 \pm 1.1)%	
3	8596 \pm 92.7	9200.4 \pm 86.3	(6.6 \pm 1.3)%	

50 < M_Z < 100 (GeV/ c^2)				
Tag Level	Measured	Predicted	$\Delta_l = (P - M)/P$	Correction
1	20571 \pm 143.4	23689.2 \pm 128.9	(13.2 \pm 0.8)%	(8.1 \pm 5.1)%
2	17203 \pm 131.2	20648.2 \pm 140.1	(16.7 \pm 0.9)%	(10.6 \pm 6.1)%
3	12299 \pm 110.9	15439.0 \pm 145.6	(20.3 \pm 1.0)%	(13.5 \pm 6.9)%

Table 6.3: Measured and Predicted Tags in the Z electron sample for a given Z mass window. The average difference is taken as a correction. The systematic error is inflated to cover the difference.

6.3 Conversion Sample Cross Check

We also examine the predictive power of the tag matrix within the conversion sample. Since we construct the tag matrix in the same sample that we test it, we expect that the average tagging efficiency should agree, but this does not imply that the tagging efficiency would be well predicted as a function of variables not in the tag matrix. We find that the agreement between the expected and measured number of tags is very good across all variables tested. We have inflated the systematic uncertainty on the tag matrix to three percent to cover some of the variation in the isolation distribution, since that is a key variable and we have not binned it finely.

Figures 6.2, 6.3, and 6.4 show the predicted and measured distributions for p_T , η , isolation, z_0 , the number of z vertices, and qd'_0 , the beamline corrected impact parameter signed by the charge of the track. Since the tag matrix is parameterized

as a function of p_T , η , and isolation, we expect the prediction and measurement to agree in these variables. We apply a 3% relative systematic uncertainty to the tag matrix prediction to cover the variation in the samples.

6.4 Calorimeter Modeling

Here we verify the claim that the variables E_{EM}/p and E_{Had}/E_{EM} are indeed well modeled in MC, particularly in terms of p_T and isolation. We avoid comparing the variables in a conversion sample to a HF electron sample directly because – even accounting for differences in the track-based isolation distribution – the environment can be different. For example, conversion electrons often originate from a photon from π^0 decay and so have a second accompanying photon that can alter the electromagnetic energy deposition. Instead, we select conversion electrons in events triggered on a 50 GeV jet. This sample is chosen to test the effect of a large hadronic component on the modeling of the calorimeter variables. A corresponding di-jet MC is used, with a filter on an outgoing parton of $p_T > 40$ GeV/ c . A jet with corrected energy > 50 GeV is required in both data and MC.

Conversions are selected by using the SLT_e to identify electrons embedded in jets. The electron identification is the same as shown in Table 4.6 plus a Level 3 SLT_e tag, except that the conversion requirements are reversed. An oppositely-charged partner track with $|\Delta \cot(\theta)| < 0.03$, $|sep| < 0.2$ cm, and $R_{conv} > 0$ cm must be found in the event. The partner must point to a different calorimeter clusters, lest the calorimeter variables are biased. To further enhance the conversion content of the sample, we also require that the SLT_e tagged leg have more than three missing silicon layers.

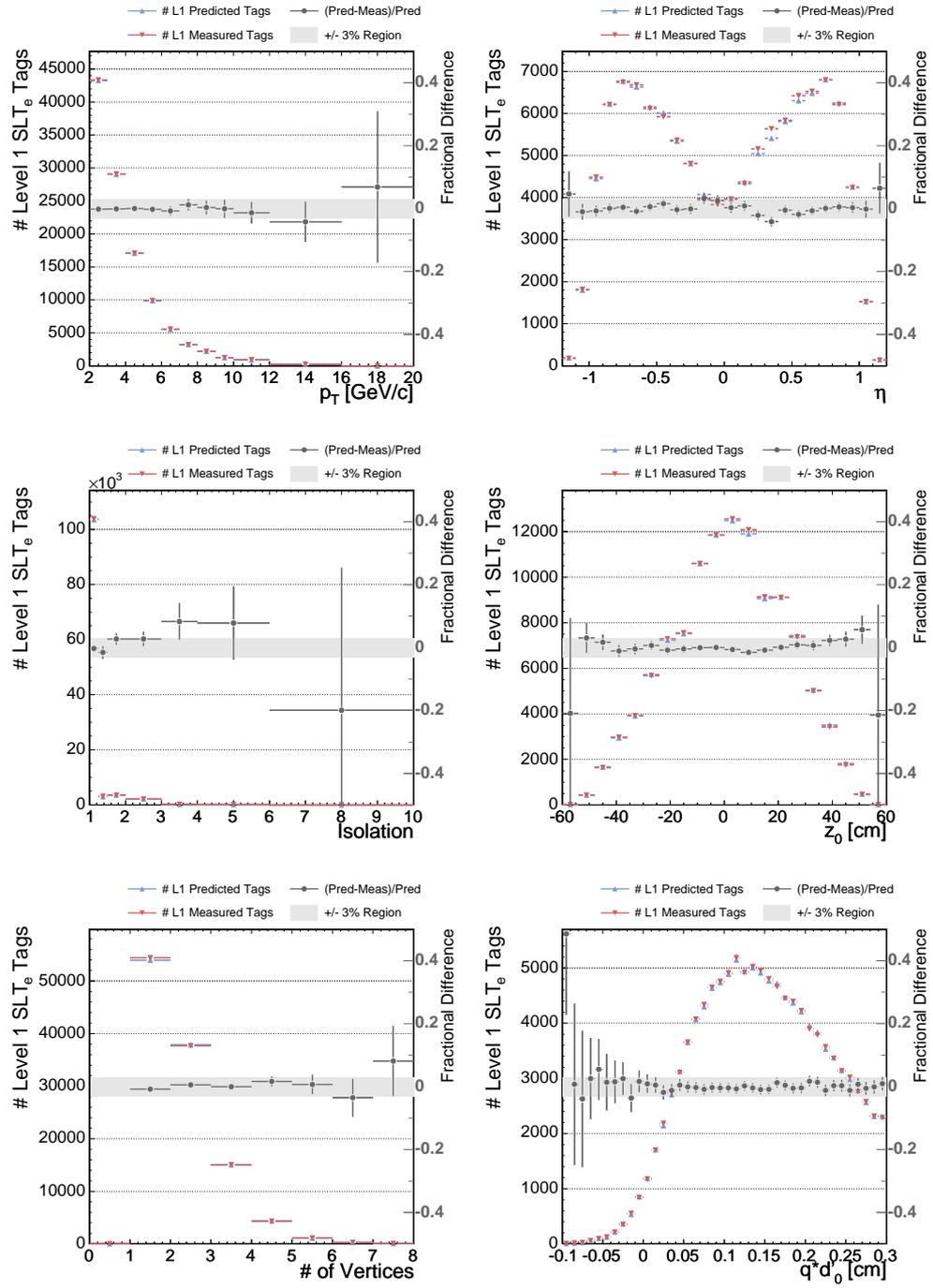


Figure 6.2: Predicted versus Measured Level 1 SLT_e Tags in the conversion electron sample.

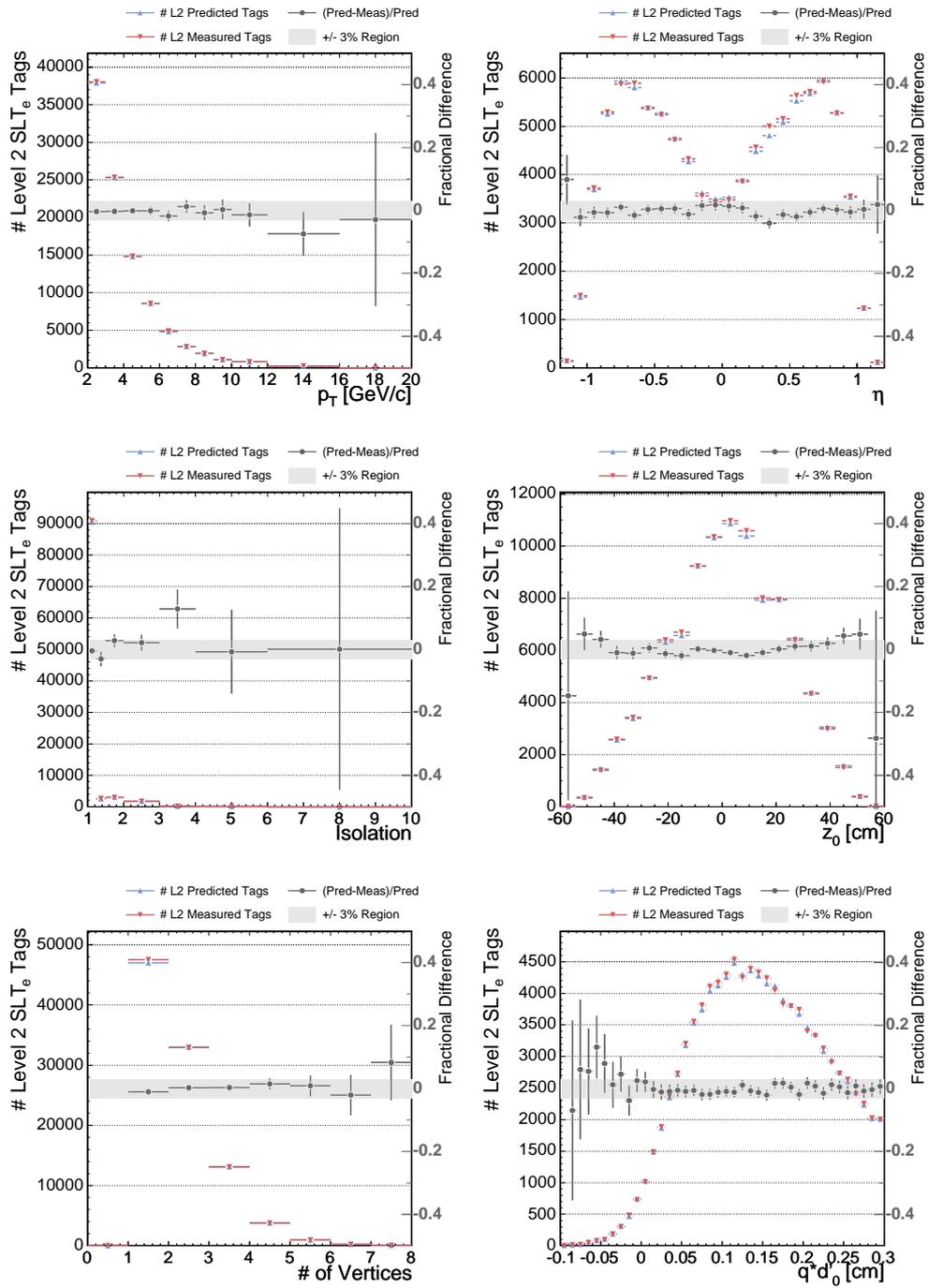


Figure 6.3: Predicted versus Measured Level 2 SLT_e Tags in the conversion electron sample.

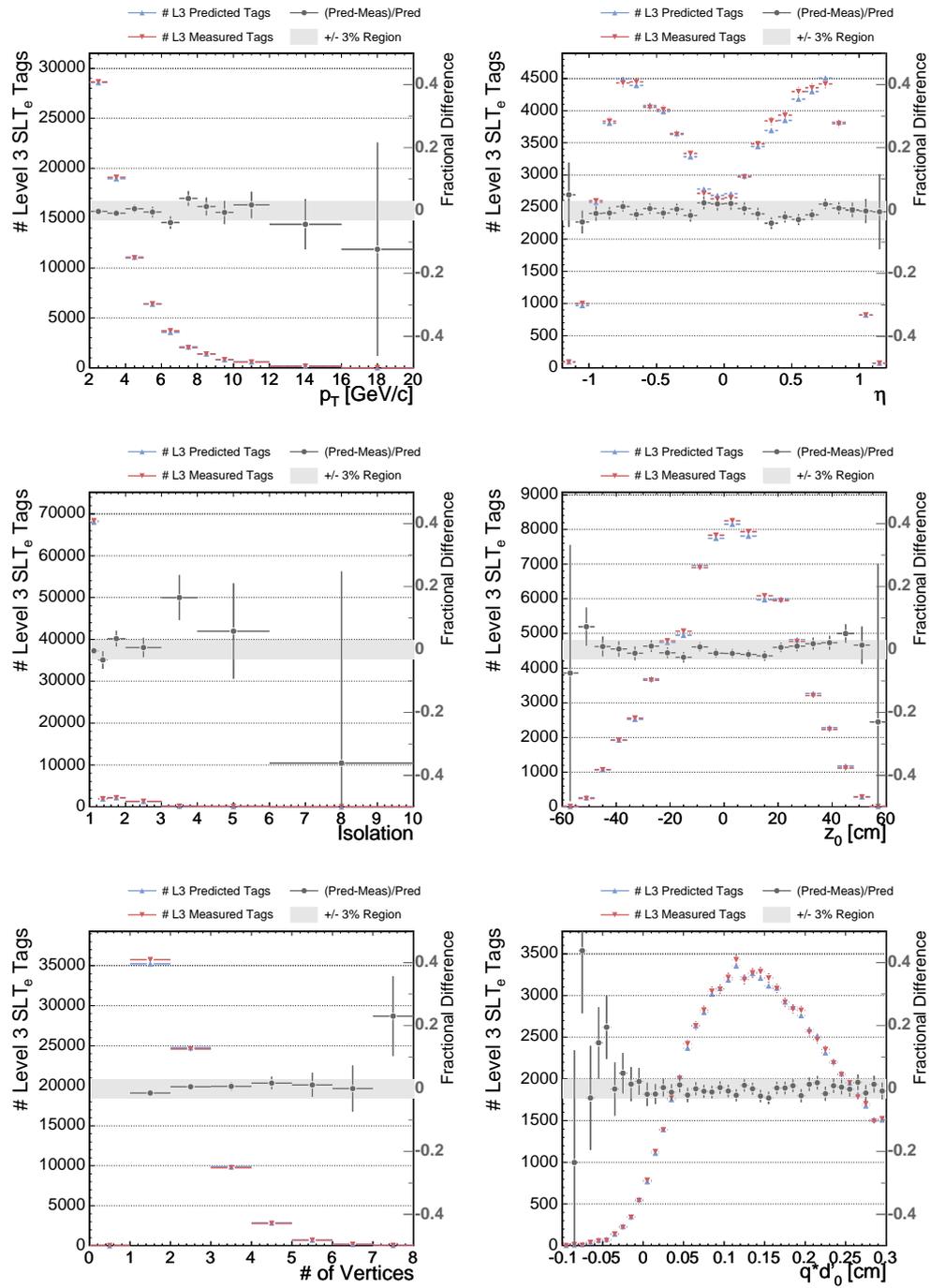


Figure 6.4: Predicted versus Measured Level 3 SLT_e Tags in the conversion electron sample.

Finally, we avoid conversion electrons matched to the highest E_T jet in the event to avoid a bias from the trigger. This selection is summarized in Table 6.4.

Conversion Selection for Calorimeter Modeling Cross Check
Event triggered on 50 GeV jet
Two tracks forming a conversion: opposite charge $ \Delta \cot(\theta) < 0.03$ $sep < 0.2$ cm $R_{conv} > 0$ cm
Point to different calorimeter clusters Not matched to highest E_T jet
One leg is taggable, passes SLT_e tag level 3, and > 3 missing SI layers Other leg used is taggable (no conversion or calorimeter requirements)

Table 6.4: Z electron sample selection summary.

Figure 6.5 compares the combined efficiency of the E_{EM}/p and E_{Had}/E_{EM} calorimeter requirements between data and MC. There is strong agreement between data and MC across p_T and in both isolated and non-isolated samples. A 2.5% relative systematic is used to cover the variation.

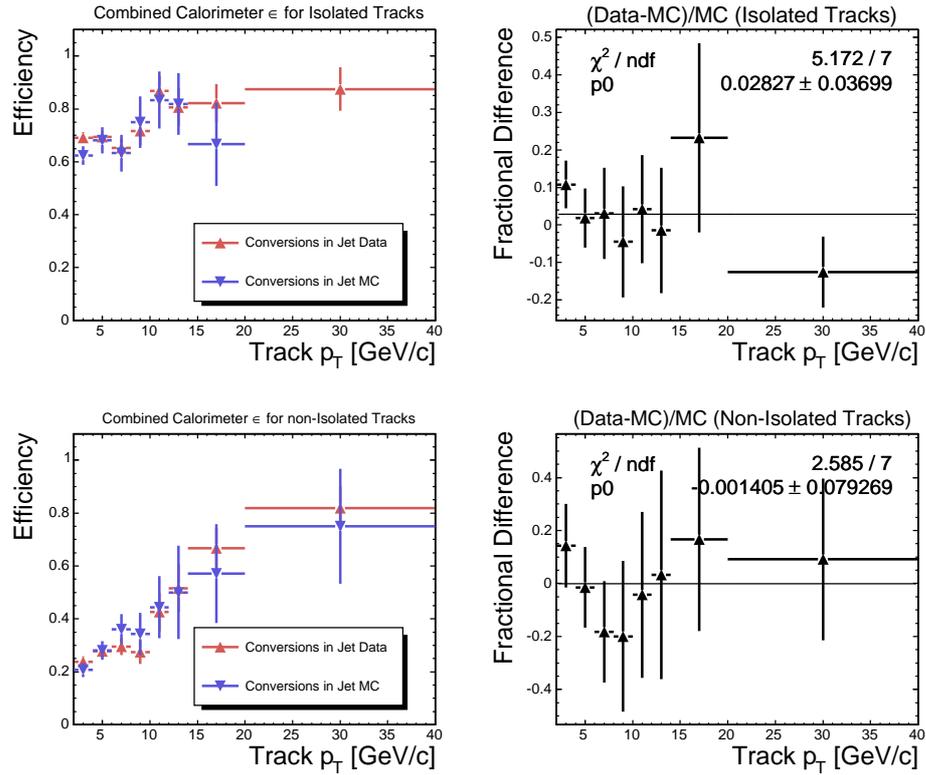


Figure 6.5: Combined E_{EM}/p and E_{Had}/E_{EM} efficiency for electrons from conversion in jet50 data (red) and MC (blue) as a function of p_T . Isolated tracks (top) are considered separately from non-isolated tracks (bottom). The difference between data and MC is also shown fitted to a constant.

Chapter 7

Fake Matrix

In this chapter, we return to our discussion on fake electron tags. We use generic tracks from events triggered on a high E_T jet as a sample of charged hadrons. In Chapter 6, we used conversion electrons to model heavy flavor (HF) electrons and built a tag matrix which predicts the tagging probability of taggable electrons in MC. By analogy, we also build a fake matrix, a parametrization of the non-electron tagging efficiency as a function of p_T , η , and isolation. Although most of the generic tracks are not electrons, there is a residual contamination of real electrons, typically originating from photon conversions or from semileptonic decay of HF. Other very small contributions include Dalitz decay of π^0 , η , and J/ψ . In Section 7.1, we define the fake matrix. In Section 7.2, we measure the real electron contribution to the generic jet datasets and correct the fake matrix appropriately. Finally, in Section 7.3, we use the corrected fake matrix to predict the tagging rate of charged pions from the decay of K_s .

7.1 Fake Matrix Definition

The fake matrix measures the efficiency

$$\varepsilon \equiv \frac{N_{tag}}{N_{taggable}}. \quad (7.1)$$

This definition of the fake matrix is identical to the tag matrix, except that we have selected generic tracks rather than conversion electrons. The definition of tag and taggable is summarized in Tables 4.7 and 4.6. Our sample is built from the generic jet samples, Jet 20, 50, 70, and 100. We avoid a trigger bias by matching a jet to the Level 2 cluster(s) which fired the trigger. We ignore tracks within the trigger jet unless another jet also exceeds the trigger threshold. This schema prevents over-sampling (or under-sampling) of jets below and above the trigger threshold.

We also reject tracks when they contain a large fraction of the total jet energy. This is done by defining the variable

$$E_{frac} \equiv \frac{Trk p_T}{Jet E_T^{corr}} \quad (7.2)$$

which is just the SLT_e track p_T divided by the corrected transverse energy of the jet. This distribution for taggable tracks from Jet 50 data and from jets in $t\bar{t}$ MC is shown in Figure 7.1. In the $t\bar{t}$ MC sample, we show the distribution for both electrons and non-electrons. As a reminder, taggable tracks are matched to a jet, have passed the conversion filter, and have an electron-like calorimeter cluster (see Table 4.6). Therefore, the taggable selection has an enhanced HF electron content relative to random tracks. Taggable electrons in $t\bar{t}$ events - primarily from HF - contain a larger fraction of the jet energy than taggable non-electrons. However, more than 90% of non-electrons and nearly 70% of electrons carry less than 25% of the energy.

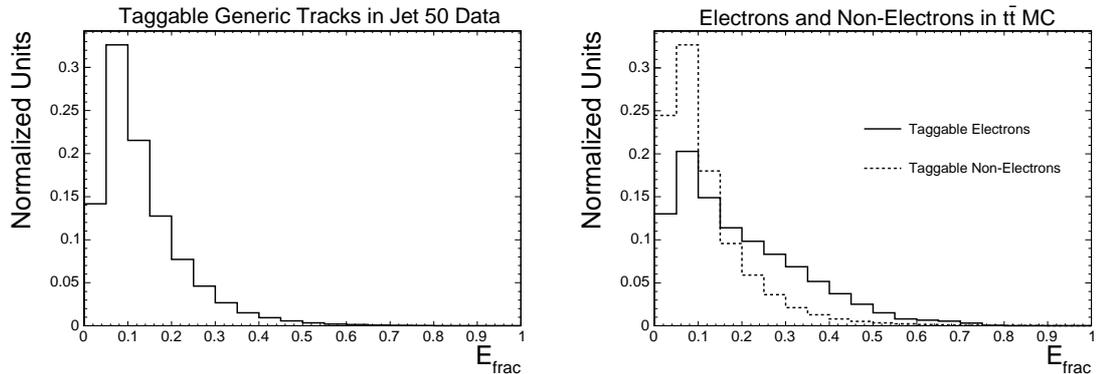


Figure 7.1: E_{frac} distribution for taggable tracks in the Jet 50 data sample (left), $t\bar{t}$ MC sample (right). The distribution on the right is divided into electrons and non-electrons.

For a track to be considered in the fake matrix, we require $E_{frac} < 0.25$. This has the effect of reducing the p_T dependence of the heavy flavor contamination of the generic jet samples. Figure 7.2 shows the fake tag rate in the four generic jet samples as a function of p_T with and without the $E_{frac} < 0.25$ requirement. Figure 7.3 shows the effect of the $E_{frac} < 0.25$ requirement on the p_T distributions of taggable tracks in each of the jet samples. Without the requirement, at high p_T the agreement between jet samples diverges. This is due to the fact that tracks containing a higher fraction of the jet energy are more likely to be HF.

For instance, a 10 GeV/ c track in the jet20 sample is more likely to be HF than a 10 GeV/ c track in the jet100 sample, simply because the former is, on average, carrying a greater fraction of the jet energy than the latter. This is demonstrated in Figure 7.4 where the E_{frac} of taggable tracks with $10 < p_T$ (GeV/ c) < 12 in both jet20 and jet100 are shown. Approximately 7.5% of these tracks in jet20 are electrons whereas only 4.5% of such tracks in jet100 are electrons, a 67% relative difference.

With the E_{frac} requirement, the agreement between the different samples is much better. In fact, the agreement is *understated* in this plot because correlations with other variables have not been taken into account. Tracks in jet20 are more isolated than tracks in the other jet samples. Once this correlation has been taken into account, the agreement between samples is very strong. Because of this, we combine all four generic jet datasets (jet20/50/70/100) together to form the fake matrix.

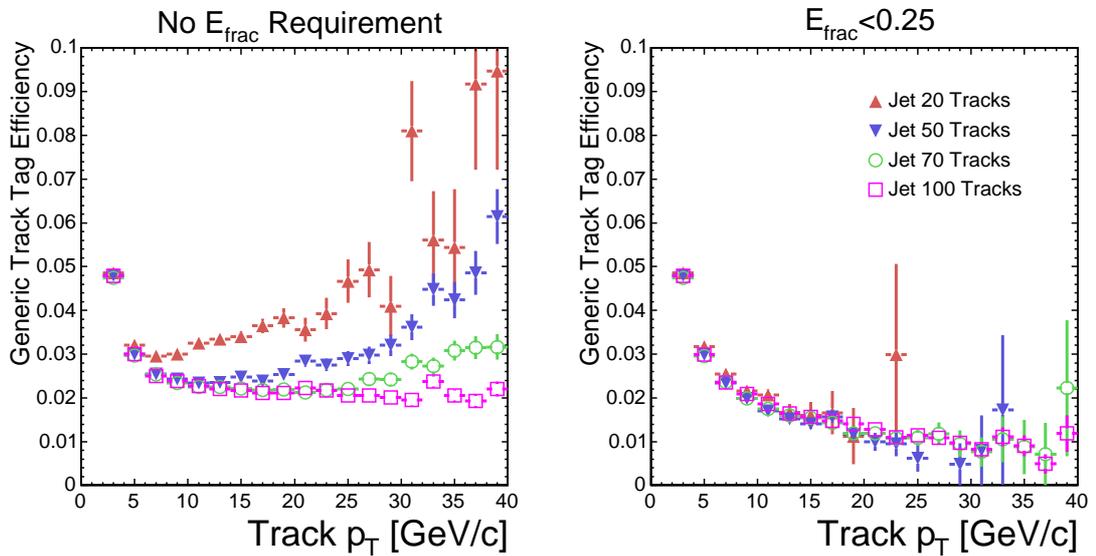


Figure 7.2: Fake electron tagging rate in different generic jet datasets as a function of p_T . SLT_e Tag Level 3 is used. No E_{frac} requirement is applied on the left, and $E_{frac} < 0.25$ is required on the right.

We bin the fake matrix as follows:

- p_T : [2.0, 2.5, 3.0, 4.0, 5.0, 7.0, 10.0, 15.0, ∞)
- $|\eta|$: [0.0, 0.4, 0.8, 1.2)
- I : $I=1.0$, $1.0 < I < 2.0$, $2.0 \leq I$

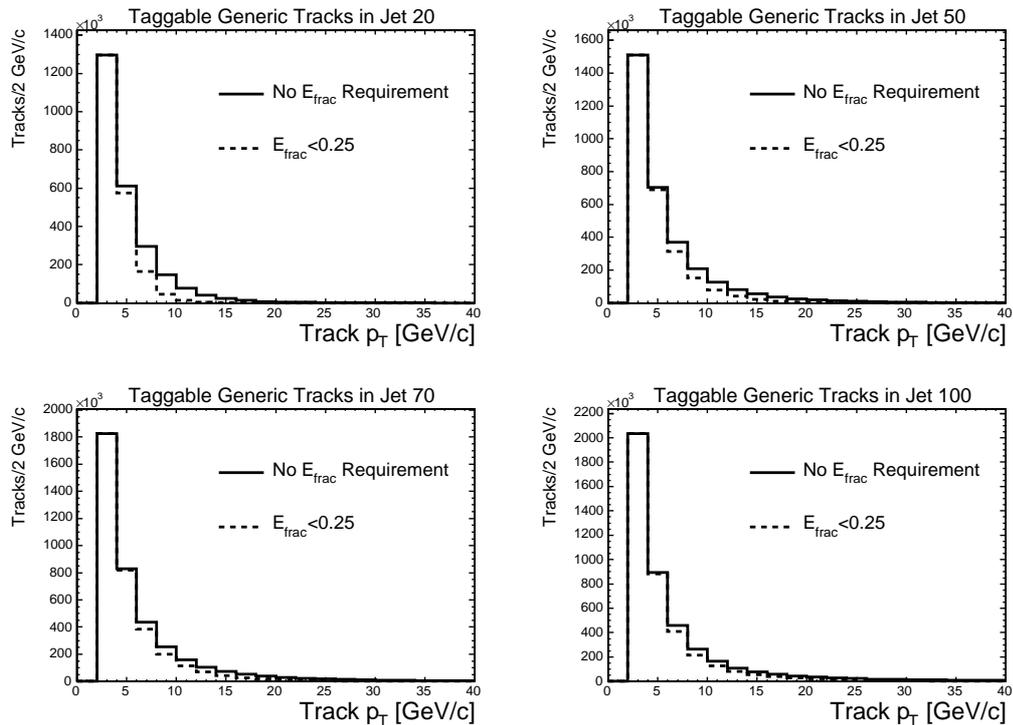


Figure 7.3: The effect of the $E_{frac} < 0.25$ requirement on the p_T distribution of taggable tracks in each of the four generic jet samples.

Unlike the tag matrix, no factorization is done, but the tagging rate varies within the η and isolation. Rather than make finer bins, we make linear corrections to the matrix.

We make the same η correction as in the tag matrix, where we correct the predicted tag rate in the forward regions ($|\eta| > 0.8$) with a linear, multiplicative correction

$$1 - a \cdot |\eta| + b. \quad (7.3)$$

We leave the central regions uncorrected. The values for the constants at each tag level are given in Table 7.1. As an example, Figure 7.5 shows the Level 1 tagging efficiency for generic tracks in Jet 50. The average efficiency is in the $|\eta| > 0.8$ region

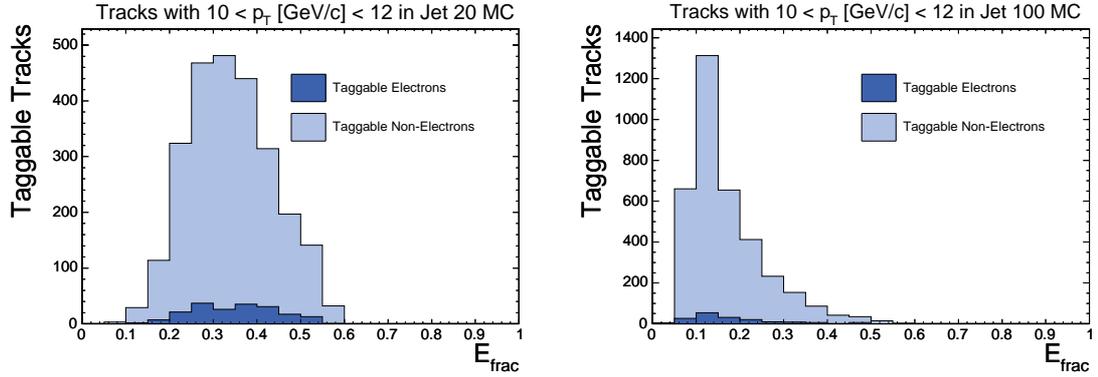


Figure 7.4: The E_{frac} distribution for taggable generic tracks with $10 < p_T \text{ (GeV/c)} < 12$ in jet20 (left) and jet100 (right) MC. The fraction of electrons in the jet20 sample is 67% higher.

is 6.6%. The lines drawn show the correction: $(1 - 0.908 \cdot |\eta| + 0.837) \cdot 0.066$.

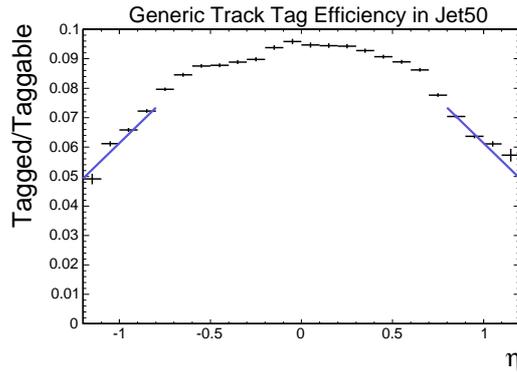


Figure 7.5: Tagging efficiency for generic tracks in the Jet 50 dataset. The lines indicate the multiplicative correction used in the fake matrix for the $|\eta| > 0.8$ region.

The fake tag rate is also a decreasing function of isolation even within the given binning. We can correct for this, as we did before with the η correction with a linear, multiplicative correction:

$$1 - a \cdot I + b. \quad (7.4)$$

We correct the isolation within the fake matrix bins $1.0 < I < 2.0$ and $2.0 < I$. The parameters are given in Table 7.2. The third set of parameters for $4 \geq I$ are chosen so that they are continuous with the $2 \leq I < 4.0$ set.

Tag Level	constant ‘a’	constant ‘b’
1	0.908	0.837
2	1.096	1.004
3	1.524	1.399

Table 7.1: Constants for the η corrections in the forward, $|\eta| > 0.8$, regions of the fake matrix.

Tag Level	constant ‘a’	constant ‘b’
$1 < \text{Isolation} < 2.0$		
1	0.32	0.49
2	0.35	0.55
3	0.41	0.64
$2 \geq \text{Isolation} < 4.0$		
1	0.13	0.36
2	0.14	0.39
3	0.16	0.45
$4 \geq \text{Isolation}$		
1	0.0	-0.16
2	0.0	-0.17
3	0.0	-0.19

Table 7.2: Constants for the isolation corrections of the fake matrix.

In Figure 7.6 we compare the isolation distribution between taggable tracks in jet50 data and MC, as well as the isolation variable’s numerator and denominator. We find the agreement to be very strong, although the agreement of its numerator,

$\sum_{clst} p_T$, and its denominator, p_T , are not as good (for a reminder of the definition of isolation, see Equation 4.1). This general agreement is important since isolation is one of the parametrization variables of the fake matrix. If the MC mis-models this variable, then the fake prediction could be incorrect.

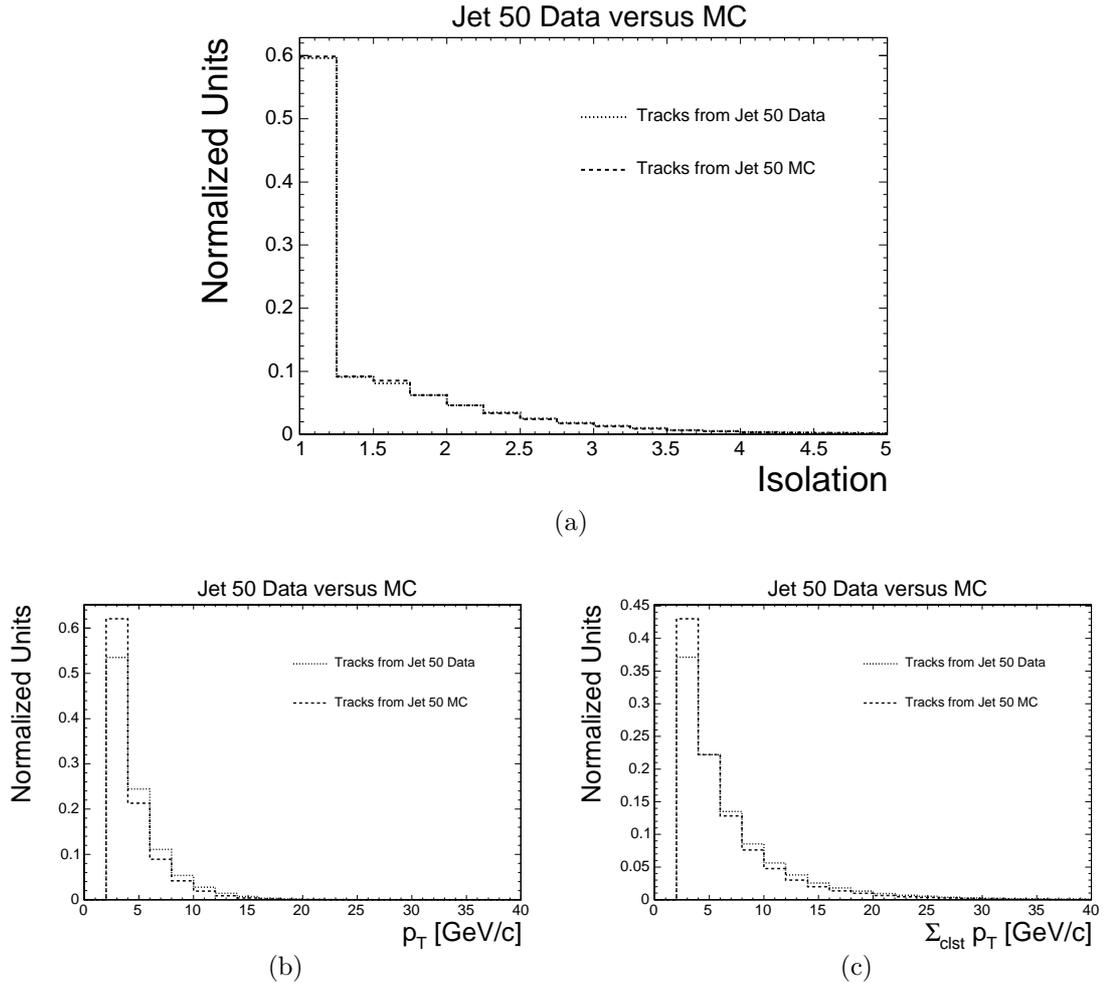


Figure 7.6: Comparison of the isolation (a), p_T (b), and $\sum_{clst} p_T$ (c) distribution of taggable tracks in jet50 data and MC.

As an initial estimate of the predictive ability of the fake matrix, we cross check it in an entirely independent sample. We compare the predicted and measured tags in

events triggered on a 20 GeV photon. Since a photon fired the trigger, and not a jet, we are not concerned with a trigger bias. We do still use the $E_{frac} < 0.25$ requirement. We find 119935 ± 346 Level 1 tags in a subset of the photon+jets sample, when we expect 114407 ± 745 tags, a 4.7% difference. The uncertainty in the tag expectation is due to the statistics in the individual matrix bins. Therefore, we apply a 5% relative systematic to the fake matrix. Figure 7.7 shows the measured and predicted tags in the photon+jet dataset.

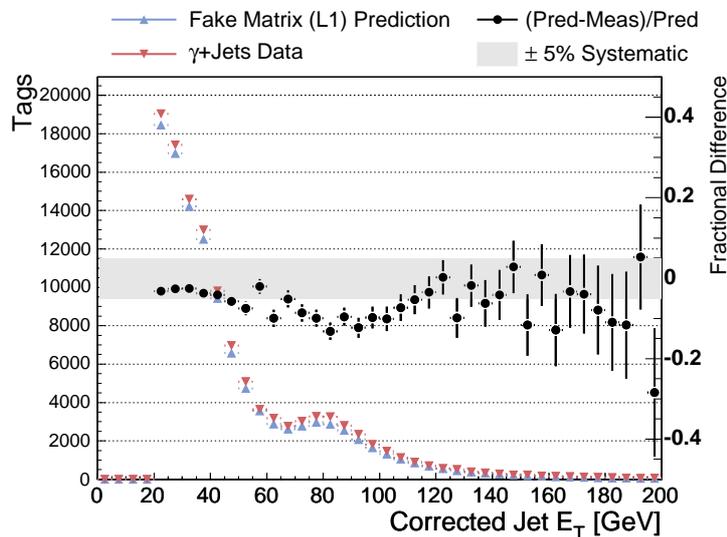


Figure 7.7: Uncorrected SLT_e L1 Fake Matrix measurement and prediction in the photon+jet sample as a function of the jet E_T .

Figures 7.8 through 7.16 show the measured and predicted tags in the different jet datasets as a function of p_T , η , isolation, z_0 , the corrected jet E_T , and the number of z vertices. The agreement between measured and predicted tags across the different variables is very good. The forward η correction is not as well predicted in the jet20 dataset, but it predicts the behavior properly in the others. The tag rate as a

function of z_0 is not well described due to variation in the conversion contamination. In addition, conversions are removed using a conversion filter with a z_0 dependent efficiency (see Figure 5.7). The presence of the 3 barrel silicon structure is evident in the difference between prediction and measurement. This will not affect the reliability of the fake matrix to predict the z_0 distribution since we do not expect fake electrons to have this material enhanced structure. The tag rate is slightly over-predicted at very high number of z vertices, probably due to local environment effects not fully captured by the isolation variable. The fake rate prediction is in terrific agreement with the measured rate as a function of the jet E_T irrespective of jet dataset. This is an important result since it demonstrates that the trigger bias is not a relevant effect within the given systematics.

Although we reject tracks with $E_{frac} < 0.25$ when we build the fake matrix, we will still want to apply the fake matrix to tracks with $E_{frac} \geq 0.25$. However real electrons contaminate the $E_{frac} \geq 0.25$ region of the generic track sample. To reduce the HF electron contribution, we can compare the predicted and measured tags in jets that are neither positively nor negatively loose SECVTX tagged but for which E_{frac} is greater than 25%. The SECVTX requirement does not fully reject HF electron tracks. In jet50, we find 1469 ± 38 tags with $E_{frac} \geq 0.25$ in jets with no SECVTX tag. When we apply the fake matrix, we predict 1033 ± 32 tags, where the uncertainty is due to the statistical uncertainty in the fake matrix bins. This means that the fake matrix underestimates the tag rate $(42 \pm 2)\%$ for fakes with $E_{frac} \geq 0.25$.

Therefore, we apply a correction to the fake matrix prediction of 1.21 ± 0.21 when it is applied to tracks with $E_{frac} \geq 0.25$. A 42% correction is almost certainly overes-

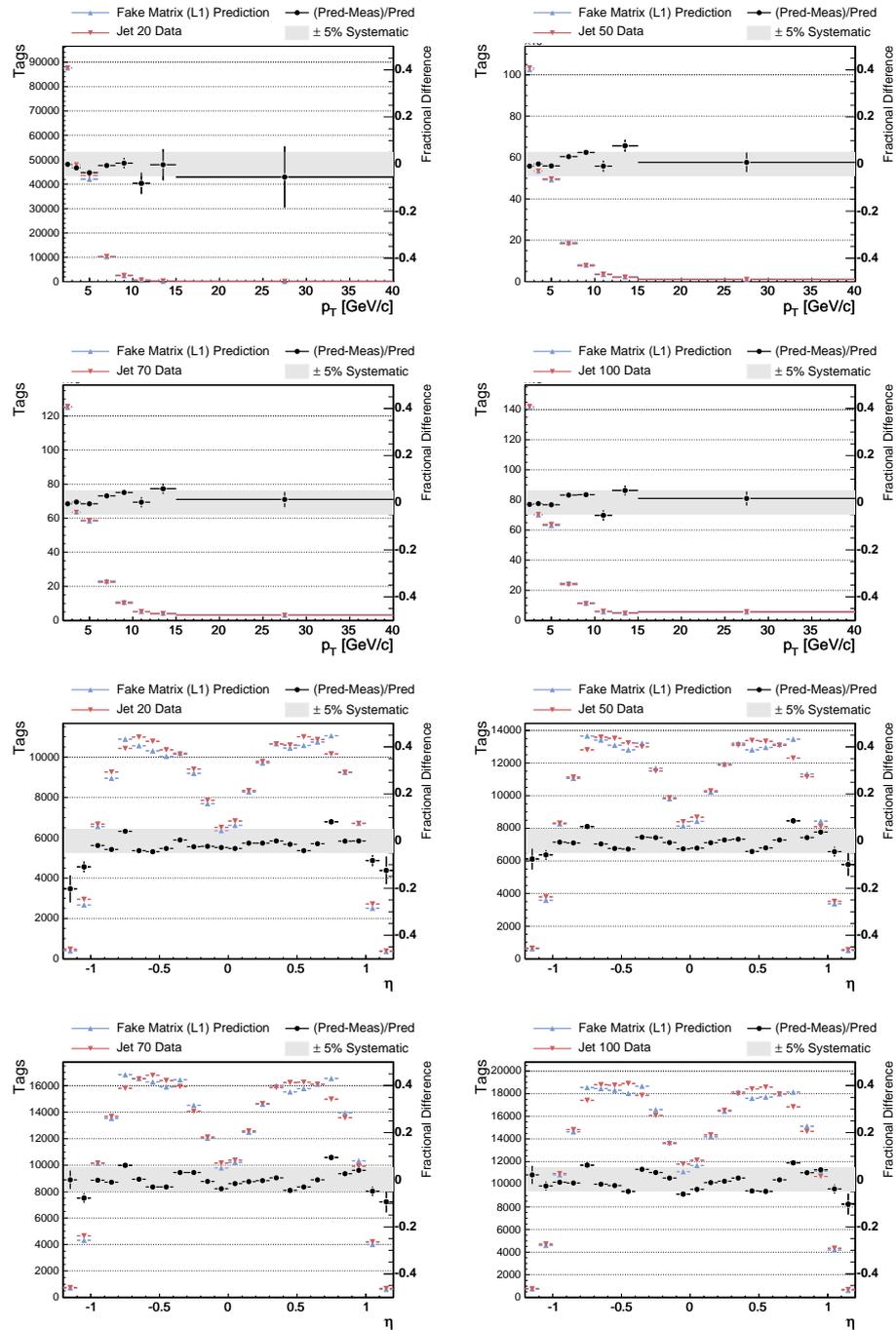


Figure 7.8: Uncorrected SLT_e L1 Fake Matrix measurement and prediction in the generic jet samples as a function of the track p_T and η .

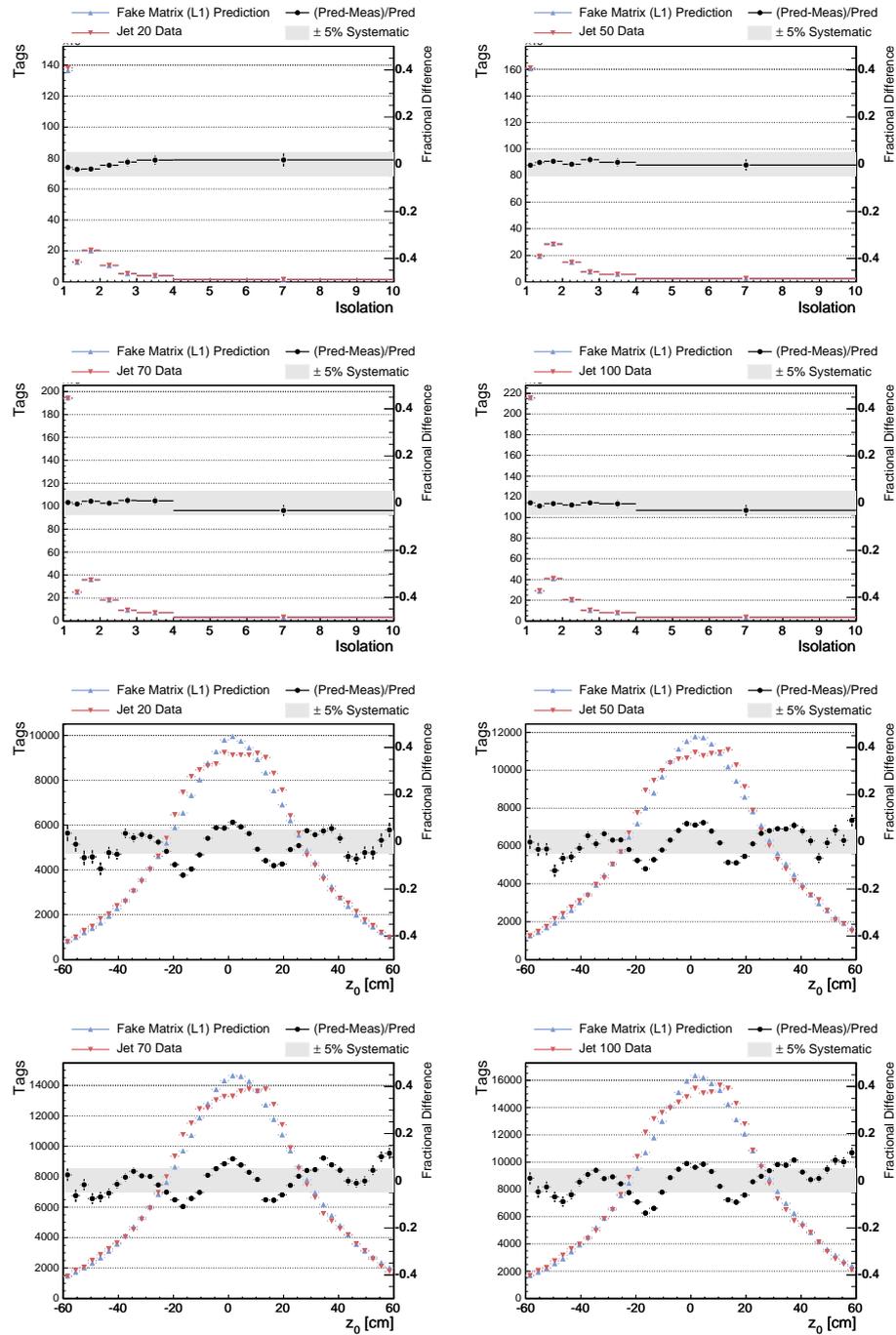


Figure 7.9: Uncorrected SLT_e L1 Fake Matrix measurement and prediction in the generic jet samples as a function of the track isolation and z_0 .

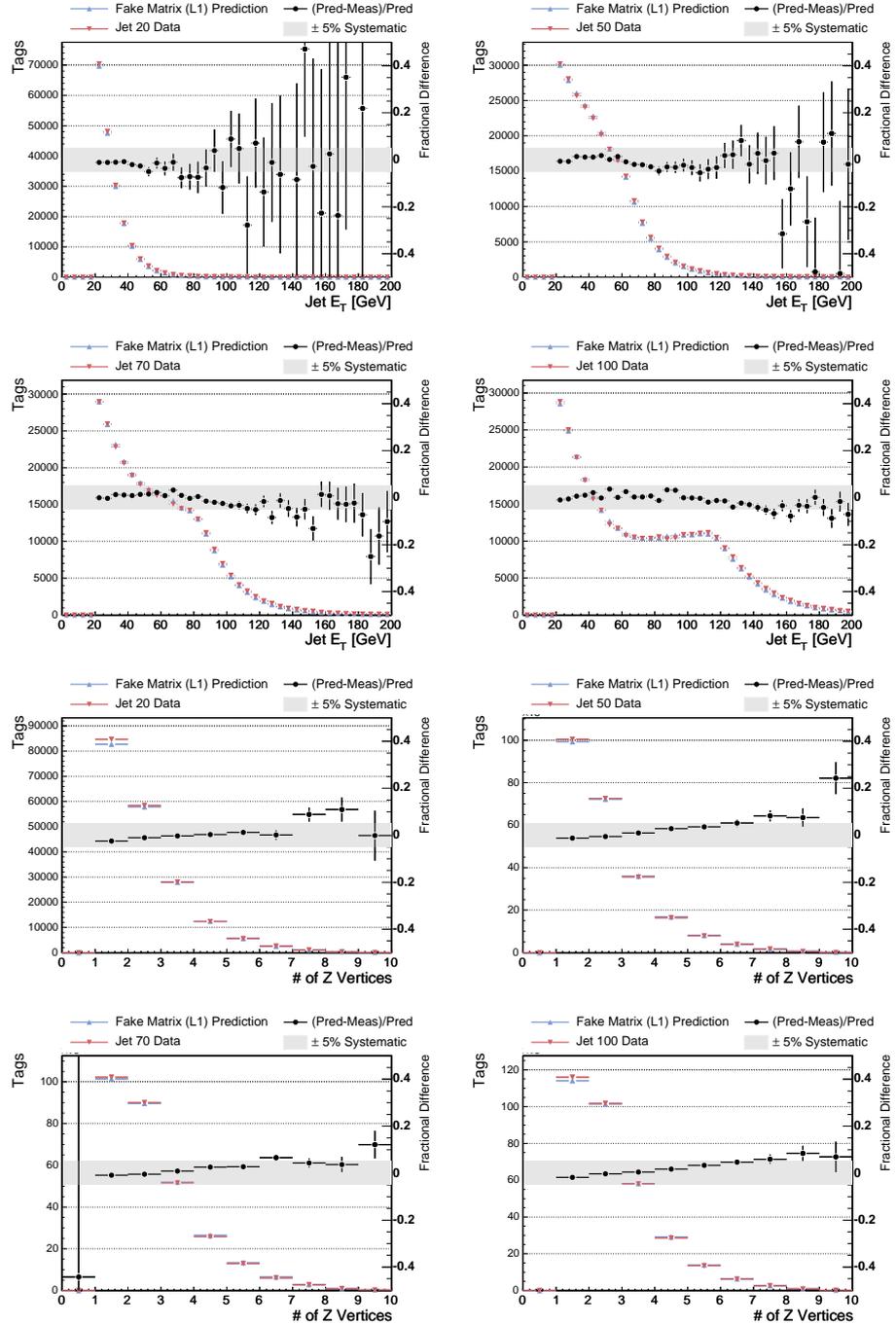


Figure 7.10: Uncorrected SLT_e L1 Fake Matrix measurement and prediction in the generic jet samples as a function of the jet E_T and the number of z vertices.

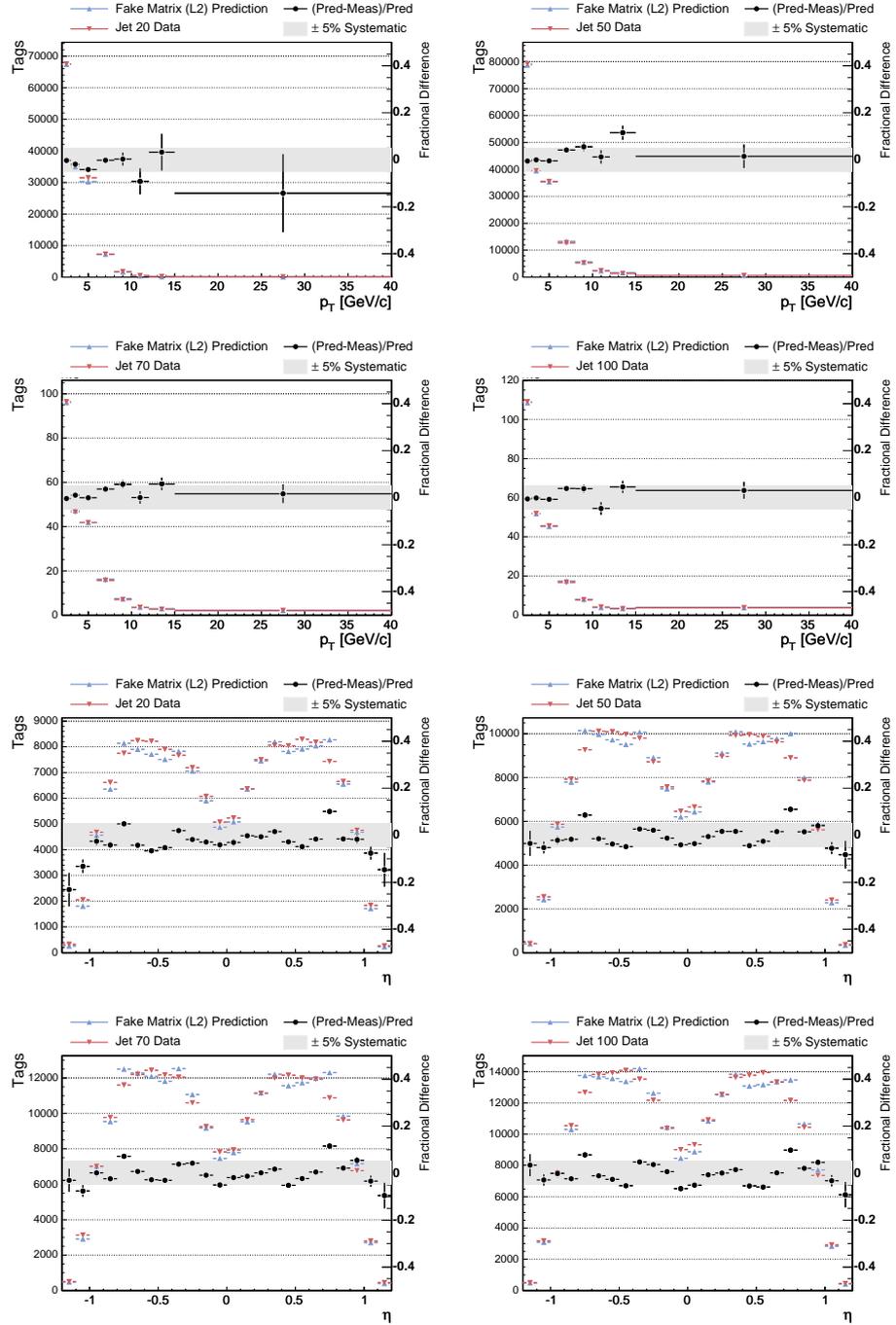


Figure 7.11: Uncorrected SLT_e L2 Fake Matrix measurement and prediction in the generic jet samples as a function of the track p_T and η .

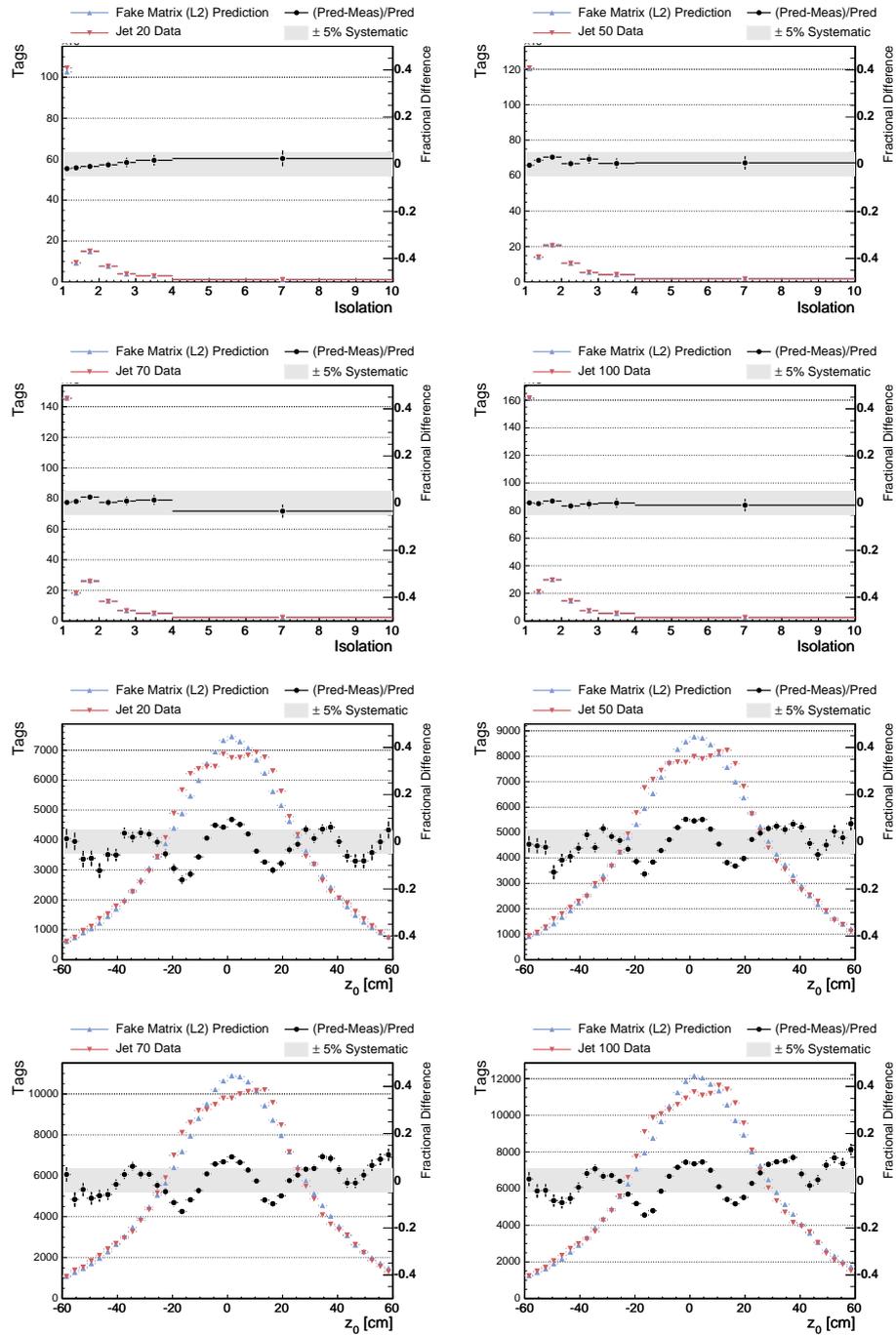


Figure 7.12: Uncorrected SLT_e L2 Fake Matrix measurement and prediction in the generic jet samples as a function of the track isolation and z_0 .

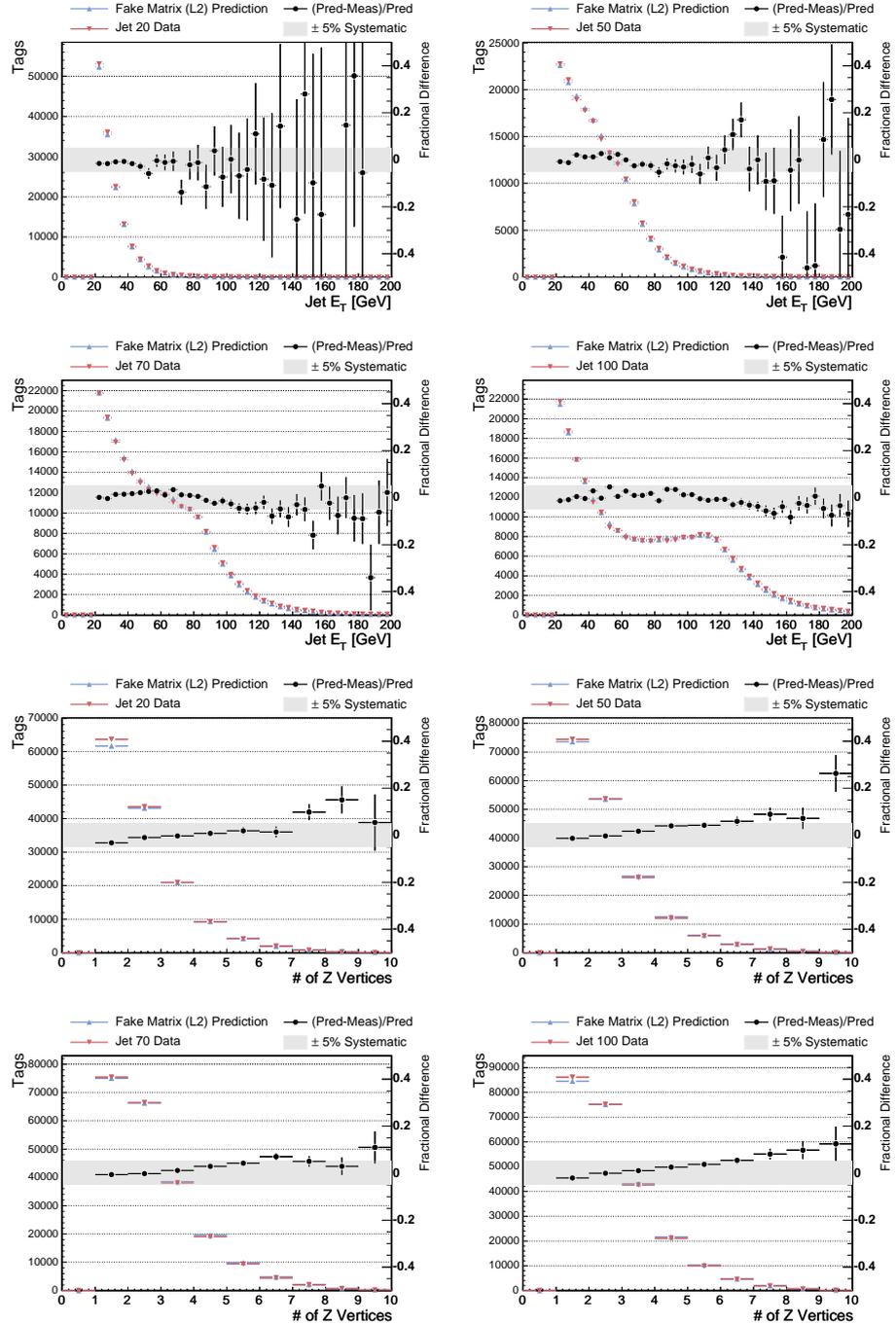


Figure 7.13: Uncorrected SLT_e L2 Fake Matrix measurement and prediction in the generic jet samples as a function of the jet E_T and the number of z vertices.

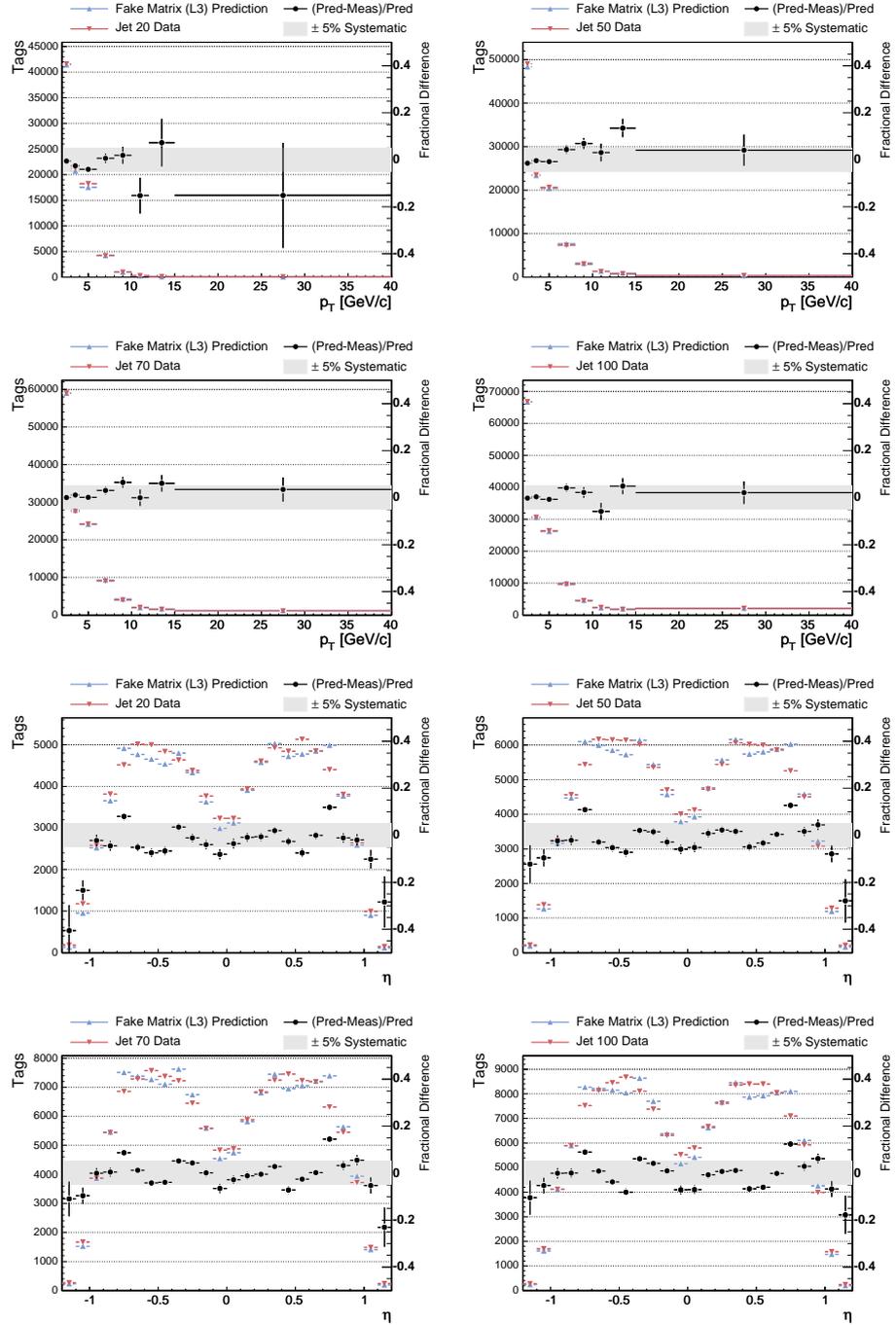


Figure 7.14: Uncorrected SLT_e L3 Fake Matrix measurement and prediction in the generic jet samples as a function of the track p_T and η .

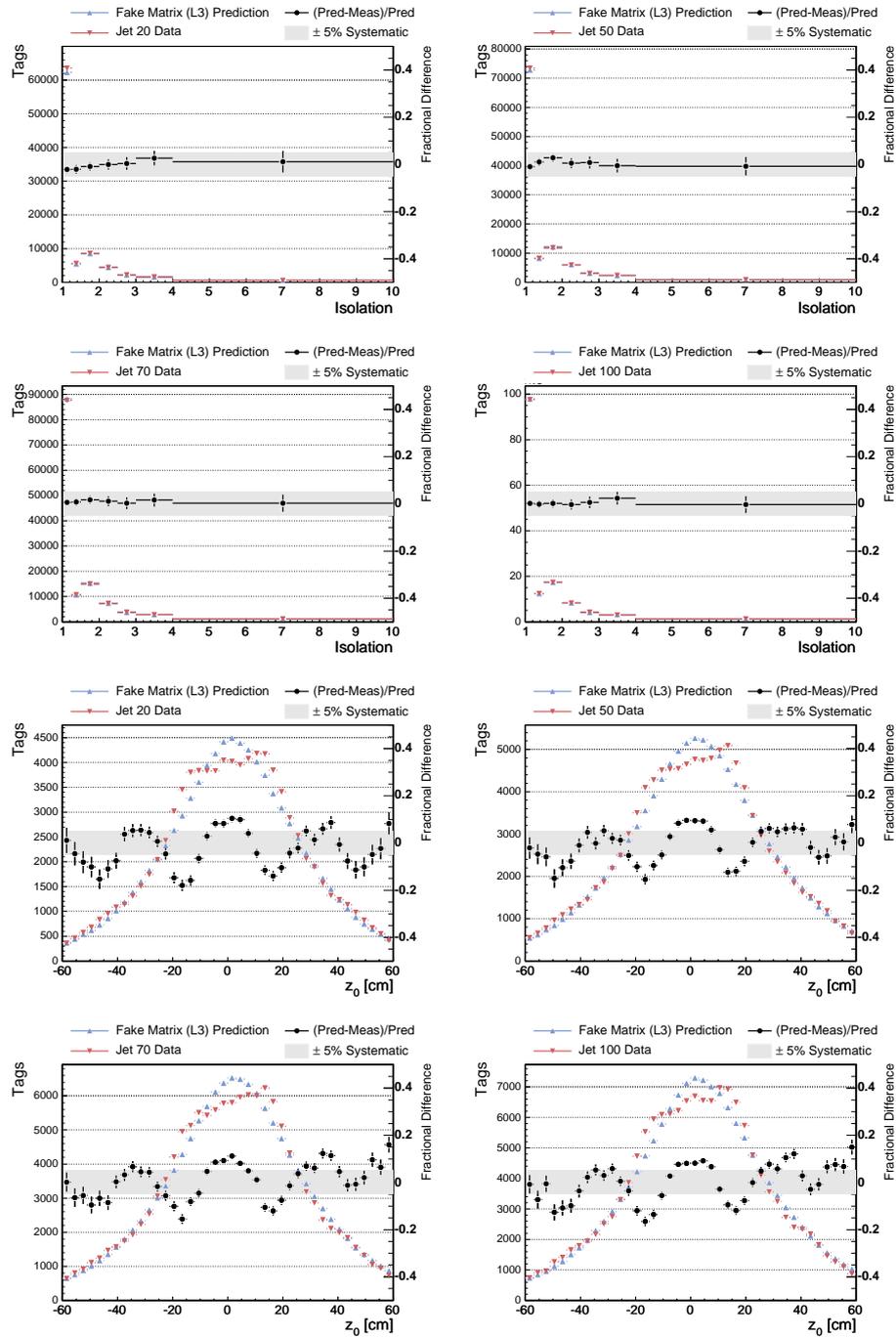


Figure 7.15: Uncorrected SLT_e L3 Fake Matrix measurement and prediction in the generic jet samples as a function of the track isolation and z_0 .

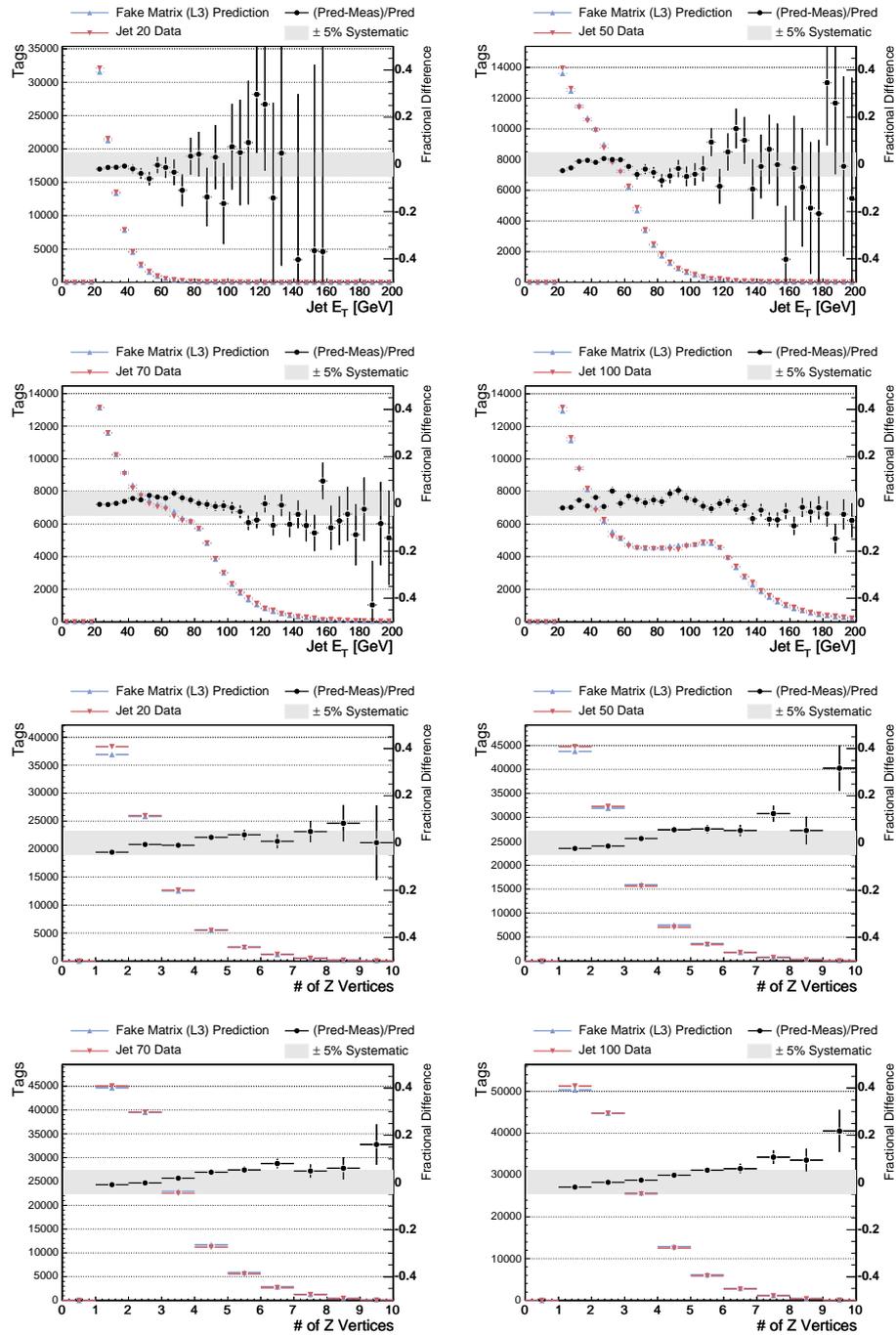


Figure 7.16: Uncorrected SLT_e L3 Fake Matrix measurement and prediction in the generic jet samples as a function of the jet E_T and the number of z vertices.

estimated since we know that HF will still be present after the SECVTX requirement. Without the SECVTX requirement, the fake matrix underestimates the tagging efficiency 57%. Since the loose SECVTX b -tagging efficiency is approximately 50%, we estimate that a 20 – 30% correction would be more appropriate.

7.2 Electron Contamination

With the $E_{frac} < 0.25$ requirement, we suppress some of the HF electron contamination of the generic jet datasets. Nonetheless, some HF electron contamination remains, along with electrons from conversions and the Dalitz decay of π^0 . In this section, we measure the total fraction of tags that are due to real electrons, F_e^{tag} , in each of the three fake matrices.

7.2.1 Conversion Electron Contamination

We determine the fraction of tags that are due to conversion electrons, F_γ^{tag} , by measuring the efficiency and over-efficiency of the conversion filter and calculating the fraction of tags that must remain due to conversions. We can divide the SLT_e tags in the jet samples between those that pass and those that fail the conversion filter. Passing the conversion filter means that the track was *not* identified as a conversion. We can also divide the SLT_e tags between those that originate from photon conversions and those that are not, calling them “prompts” as we did in the Chapter 5. This is expressed in the equation

$$N = N_{pass} + N_{fail} = N_\gamma + N_p \quad (7.5)$$

where N is the number of SLT_e tags in the jet samples before conversion removal. We can also calculate the number of SLT_e tags that fail the conversion filter with the equation

$$N_{fail} = N_\gamma \varepsilon + N_p \varepsilon^o \quad (7.6)$$

where ε and ε^o is the efficiency and over-efficiency of the filter, respectively. F_γ^{tag} is determined by the equation:

$$F_\gamma^{tag} = \frac{N_\gamma(1 - \varepsilon)}{N_{pass}} \quad (7.7)$$

since it is just the ratio of tags originating from conversions to total tags after the conversion filter has been applied. A little algebraic manipulation reveals that

$$N_{fail} = N - N_{pass} = N_\gamma \varepsilon + (N - N_\gamma) \varepsilon^o \quad (7.8)$$

so that

$$N_\gamma = \frac{N(1 - \varepsilon^o) - N_{pass}}{\varepsilon - \varepsilon^o}. \quad (7.9)$$

Therefore,

$$F_\gamma^{tag} = \left(\frac{1 - \varepsilon^o}{F_{pass}} - 1 \right) \left(\frac{1 - \varepsilon}{\varepsilon - \varepsilon^o} \right) \quad (7.10)$$

where F_{pass} is the fraction of tags that pass the conversion filter, that is, $F_{pass} \equiv N_{pass}/N$.

We measure the efficiency and over-efficiency of the conversion filter using the techniques developed in Section 5.5. The numbers here are different than those shown in Table 5.7 because of the E_{frac} requirement. Table 7.3 shows the fraction of tags in the generic jet datasets due to conversions for each of the tag levels. We can see that the fractions increase with tighter SLT_e operating points, as we would expect. It is also interesting to note the agreement within a few percent in F_γ^{tag} between datasets.

Dataset	ε (%)	ε^o (%)	F_{pass}	F_{γ}^{tag}
SLT_e Tag Level 1				
jet20	67.8 ± 1.2 ± 1.3	6.0 ± 1.6	0.617 ± 0.001	0.273 ± 0.016 ± 0.019
jet50	67.2 ± 1.5 ± 1.4	7.8 ± 2.1	0.620 ± 0.001	0.269 ± 0.020 ± 0.020
jet70	68.5 ± 1.6 ± 1.5	8.9 ± 2.4	0.619 ± 0.001	0.249 ± 0.018 ± 0.020
jet100	68.2 ± 1.6 ± 1.5	10.3 ± 2.8	0.586 ± 0.001	0.247 ± 0.020 ± 0.022
SLT_e Tag Level 2				
jet20	67.8 ± 1.2 ± 1.3	6.0 ± 1.6	0.594 ± 0.001	0.303 ± 0.018 ± 0.020
jet50	67.2 ± 1.5 ± 1.4	7.8 ± 2.1	0.599 ± 0.001	0.299 ± 0.022 ± 0.022
jet70	68.5 ± 1.6 ± 1.5	8.9 ± 2.4	0.698 ± 0.001	0.276 ± 0.021 ± 0.022
jet100	68.2 ± 1.6 ± 1.5	10.3 ± 2.8	0.586 ± 0.001	0.274 ± 0.022 ± 0.024
SLT_e Tag Level 3				
jet20	67.8 ± 1.2 ± 1.3	6.0 ± 1.6	0.561 ± 0.001	0.352 ± 0.021 ± 0.023
jet50	67.2 ± 1.5 ± 1.4	7.8 ± 2.1	0.570 ± 0.001	0.341 ± 0.025 ± 0.025
jet70	68.5 ± 1.6 ± 1.5	8.9 ± 2.4	0.568 ± 0.001	0.318 ± 0.024 ± 0.024
jet100	68.2 ± 1.6 ± 1.5	10.3 ± 2.8	0.571 ± 0.001	0.314 ± 0.025 ± 0.026

Table 7.3: Fraction of SLT_e tags that originate from conversions given by dataset and SLT_e tag level. The uncertainties shown for the efficiency and F_{γ}^{tag} are statistical followed by systematic. The uncertainties shown for the over-efficiency, ε^o , are systematic only (the statistical uncertainties are negligible).

7.2.2 Heavy Flavor Electron Contamination

We calculate the fraction of tags in the generic jets originating from the semileptonic decay of HF, F_{HF}^{tag} , by using correlations with the SECVTX tagger. Let N^{SLT} be the number of SLT_e tags in a given sample, and let $N^{SLT+SVX}$ be the number of SLT_e tags that are also in positively tight SECVTX tagged jets. We can divide the doubly tagged tracks into those that originate from the semi-leptonic decay of HF and those that do not, that is:

$$N^{SLT+SVX} = N_{HF}^{SLT+SVX} + N_{LF}^{SLT+SVX}. \quad (7.11)$$

The designation LF applies to all tracks, including real electrons, that do not originate from semi-leptonic decay of HF. This includes, for example, conversion electrons embedded in a b -jet. The number of doubly tagged tracks from HF is simply

$$N_{HF}^{SLT+SVX} = N^{SLT} F_{HF}^{tag} \cdot \varepsilon_{SVX} \quad (7.12)$$

where ε_{SVX} is the efficiency for SECVTX to tag a jet with a HF electron inside of it. Note that F_{HF}^{tag} is the quantity of interest.

To calculate the number of doubly tagged tracks from LF, we appeal to the electron contaminated fake matrix, M_{ijk} , where the indices represent the p_T , η , and isolation bins. If the fake matrix is constructed within the same sample it is applied, then

$$N^{SLT} = N_{ijk} \cdot M_{ijk} \quad (7.13)$$

is true by construction, where we defined N_{ijk} as the total number of tracks with a given p_T , η , and isolation, and there is an implied sum over indices. To estimate the amount of LF tags, we can subtract out the HF component and again apply it to the sample:

$$N_{LF}^{SLT} = N_{ijk} \cdot M_{ijk} (1 - F_{HF}^{tag}). \quad (7.14)$$

Therefore, we can derive the equation:

$$N_{LF}^{SLT+SVX} = N_{ijk}^{SVX} \cdot M_{ijk} (1 - F_{HF}^{tag}). \quad (7.15)$$

where N^{SVX} is the number of tracks embedded in SECVTX tagged jets. The assumption here is that SECVTX tagging a jet does not alter the LF components of the SLT_e . Combining Equations 7.11, 7.12, and 7.15, we get the result:

$$F_{HF}^{tag} = \frac{N^{SLT+SVX} - N_{ijk}^{SVX} \cdot M_{ijk}}{N^{SLT} \varepsilon_{SVX} - N_{ijk}^{SVX} \cdot M_{ijk}} \quad (7.16)$$

To determine the efficiency, we appeal to MC, so that $\varepsilon_{SVX} = \varepsilon_{SVX}^{MC} \cdot SF$, where SF is the SECVTX Scale Factor (see Section 3.4.2). All of the other quantities in equation 7.16 can be measured directly in the dataset in question. Table 7.4 shows the values of F_{HF}^{tag} calculated for different jet samples and tag levels. The fraction of tags from HF varies from 5 – 9% depending on the jet dataset and tag level.

Dataset	N^{SLT}	$N^{SLT+SVX}$	$N_{ijk}^{SVX} \cdot M_{ijk}$	$\varepsilon_{SVX}^{MC} \cdot SF$	F_{HF}^{tag}
SLT_e Tag Level 1					
jet20	193116	6965	4810. ± 30.	0.24 ± 0.01 ± 0.01	0.053 ± 0.004 ± 0.003
jet50	239221	13518	9694. ± 68.	0.31 ± 0.01 ± 0.02	0.059 ± 0.003 ± 0.004
jet70	292556	19244	14164. ± 103.	0.34 ± 0.01 ± 0.02	0.060 ± 0.003 ± 0.004
jet100	327677	24264	18377. ± 137.	0.33 ± 0.01 ± 0.02	0.066 ± 0.004 ± 0.004
SLT_e Tag Level 2					
jet20	144405	5449	3572. ± 26.	0.24 ± 0.01 ± 0.01	0.062 ± 0.004 ± 0.004
jet50	176796	10395	7149. ± 58.	0.31 ± 0.01 ± 0.02	0.068 ± 0.003 ± 0.004
jet70	215837	14673	10426. ± 88.	0.34 ± 0.01 ± 0.02	0.068 ± 0.004 ± 0.004
jet100	242193	18500	13532. ± 117.	0.33 ± 0.01 ± 0.02	0.075 ± 0.005 ± 0.005
SLT_e Tag Level 3					
jet20	86671	3534	2134. ± 20.	0.24 ± 0.01 ± 0.01	0.077 ± 0.006 ± 0.005
jet50	105937	6676	4252. ± 45.	0.31 ± 0.01 ± 0.02	0.084 ± 0.004 ± 0.005
jet70	128333	9232	6192. ± 68.	0.34 ± 0.01 ± 0.02	0.082 ± 0.005 ± 0.005
jet100	144186	11496	8043. ± 91.	0.33 ± 0.01 ± 0.02	0.088 ± 0.006 ± 0.006

Table 7.4: Fraction of SLT_e tags that originate from the semi-leptonic decay of HF across dataset and SLT_e tag level.

7.2.3 Dalitz and Other Electron Contamination

There is a non-negligible contribution of real electrons in the fake matrix due neither to HF electrons nor conversion electrons. $\pi^0 \rightarrow e^+e^-\gamma$ is the dominant process not already account for. A very small contribution of $\eta \rightarrow e^+e^-$ and $J/\psi \rightarrow e^+e^-$ also exists. We can rely on MC to determine the relative amount of these sources of

electrons to HF electrons and conversions.

We measure the fraction of tags in generic jets due to Dalitz decay and other sources, F_D^{tag} , by measuring the ratio of HF electrons to other electron sources in MC, N_{HF}^{MC}/N_D^{MC} . We assume that this ratio holds in data as well, so that:

$$F_D^{tag} = F_{HF}^{tag} \frac{N_D^{MC}}{N_{HF}^{MC}}. \quad (7.17)$$

As a validation of this technique, we can compare the ratio $F_{HF}^{tag}/F_\gamma^{tag}$ to $N_{HF}^{MC}/N_\gamma^{MC}$.

In the worst instance of this, we find

$$\frac{F_{HF}^{tag}}{F_\gamma^{tag}} = 0.238 \pm 0.021 \text{ (stat)} \pm 0.034 \text{ (syst)} \quad (7.18)$$

in jet50 data and with tag level 1, versus

$$\frac{N_{HF}^{MC}}{N_\gamma^{MC}} = 0.314 \pm 0.006 \text{ (stat)} \quad (7.19)$$

in jet50 MC. The reason that the agreement is not better is because we have measured the pretag ratios in MC and compared them to the tag ratios in data. The data-MC efficiency SF has also not been applied. Therefore, we apply a 25% relative systematic to our measurement of F_D^{tag} with this technique. Table 7.5 shows the results for F_D^{tag} in the different datasets and for different tag levels.

7.2.4 Total Electron Contamination

Table 7.6 shows the total fraction of SLT_e tags in the generic jet datasets that are real electrons, F_e^{tag} . We take the direct sum of F_γ^{tag} , F_{HF}^{tag} , and F_D^{tag} to make the calculation. The agreement between jet samples is remarkable. Although the less energetic jet samples seem to have more conversion electrons, the more energetic jet

Dataset	F_{HF}^{tag}	N_{HF}^{MC}/N_D^{MC}	F_D^{tag}
SLT_e Tag Level 1			
jet20	$0.053 \pm 0.004 \pm 0.003$	2.27 ± 0.12	$0.023 \pm 0.002 \pm 0.006$
jet50	$0.059 \pm 0.003 \pm 0.004$	2.63 ± 0.08	$0.022 \pm 0.001 \pm 0.006$
jet70	$0.060 \pm 0.003 \pm 0.004$	2.40 ± 0.13	$0.025 \pm 0.002 \pm 0.006$
jet100	$0.066 \pm 0.004 \pm 0.004$	1.90 ± 0.10	$0.035 \pm 0.003 \pm 0.009$
SLT_e Tag Level 2			
jet20	$0.062 \pm 0.004 \pm 0.004$	2.27 ± 0.12	$0.027 \pm 0.002 \pm 0.007$
jet50	$0.068 \pm 0.003 \pm 0.004$	2.63 ± 0.08	$0.026 \pm 0.001 \pm 0.007$
jet70	$0.068 \pm 0.004 \pm 0.004$	2.40 ± 0.13	$0.028 \pm 0.002 \pm 0.007$
jet100	$0.075 \pm 0.005 \pm 0.005$	1.90 ± 0.10	$0.040 \pm 0.003 \pm 0.010$
SLT_e Tag Level 3			
jet20	$0.077 \pm 0.006 \pm 0.005$	2.27 ± 0.12	$0.034 \pm 0.003 \pm 0.009$
jet50	$0.084 \pm 0.004 \pm 0.005$	2.63 ± 0.08	$0.032 \pm 0.002 \pm 0.008$
jet70	$0.082 \pm 0.005 \pm 0.005$	2.40 ± 0.13	$0.034 \pm 0.003 \pm 0.009$
jet100	$0.088 \pm 0.006 \pm 0.006$	1.90 ± 0.10	$0.046 \pm 0.004 \pm 0.012$

Table 7.5: Fraction of SLT_e tags that originate from other sources of electrons across dataset and SLT_e tag level.

samples have more HF electrons, making up the difference. In any case, the trend is hardly statistically significant. Therefore, we combine all four jet samples together, and measure a combined F_e^{tag} by taking a weighted average. F_e^{tag} varies from 35% to 45% depending on the tag level.

Although we have calculated the total electron contamination in the fake matrices, we cannot blindly apply an overall correction to every bin as some bins may have more electrons than others. Figure 7.17 shows the weighted average over all four jet samples of F_e^{tag} by p_T , η , and isolation bin in the fake matrix. We show the tag level 1 matrix since it has the smallest uncertainties. Keep in mind that because we are fitting shapes for the value of F_γ^{tag} in individual bins, the mean of the fits may not equal the overall fit.

Dataset	F_γ^{tag}	F_{HF}^{tag}	F_D^{tag}	F_e^{tag}
SLT_e Tag Level 1				
jet20	$0.27 \pm 0.02 \pm 0.02$	$0.05 \pm 0.00 \pm 0.00$	$0.02 \pm 0.00 \pm 0.01$	$0.35 \pm 0.02 \pm 0.02$
jet50	$0.27 \pm 0.02 \pm 0.02$	$0.06 \pm 0.00 \pm 0.00$	$0.02 \pm 0.00 \pm 0.01$	$0.35 \pm 0.02 \pm 0.02$
jet70	$0.25 \pm 0.02 \pm 0.02$	$0.06 \pm 0.00 \pm 0.00$	$0.03 \pm 0.00 \pm 0.01$	$0.33 \pm 0.02 \pm 0.02$
jet100	$0.25 \pm 0.02 \pm 0.02$	$0.07 \pm 0.00 \pm 0.00$	$0.04 \pm 0.00 \pm 0.01$	$0.35 \pm 0.02 \pm 0.02$
comb.	—	—	—	$0.35 \pm 0.01 \pm 0.02$
SLT_e Tag Level 2				
jet20	$0.30 \pm 0.02 \pm 0.02$	$0.06 \pm 0.00 \pm 0.00$	$0.03 \pm 0.00 \pm 0.01$	$0.39 \pm 0.02 \pm 0.02$
jet50	$0.30 \pm 0.02 \pm 0.02$	$0.07 \pm 0.00 \pm 0.00$	$0.03 \pm 0.00 \pm 0.01$	$0.39 \pm 0.02 \pm 0.02$
jet70	$0.28 \pm 0.02 \pm 0.02$	$0.07 \pm 0.00 \pm 0.00$	$0.03 \pm 0.00 \pm 0.01$	$0.37 \pm 0.02 \pm 0.02$
jet100	$0.27 \pm 0.02 \pm 0.02$	$0.08 \pm 0.00 \pm 0.00$	$0.04 \pm 0.00 \pm 0.01$	$0.39 \pm 0.02 \pm 0.03$
comb.	—	—	—	$0.39 \pm 0.01 \pm 0.02$
SLT_e Tag Level 3				
jet20	$0.35 \pm 0.02 \pm 0.02$	$0.08 \pm 0.01 \pm 0.00$	$0.03 \pm 0.00 \pm 0.01$	$0.46 \pm 0.02 \pm 0.03$
jet50	$0.34 \pm 0.03 \pm 0.03$	$0.08 \pm 0.00 \pm 0.01$	$0.03 \pm 0.00 \pm 0.01$	$0.46 \pm 0.03 \pm 0.03$
jet70	$0.32 \pm 0.02 \pm 0.02$	$0.08 \pm 0.00 \pm 0.01$	$0.03 \pm 0.00 \pm 0.01$	$0.44 \pm 0.03 \pm 0.03$
jet100	$0.31 \pm 0.03 \pm 0.03$	$0.09 \pm 0.01 \pm 0.01$	$0.05 \pm 0.00 \pm 0.01$	$0.45 \pm 0.03 \pm 0.03$
comb.	—	—	—	$0.45 \pm 0.01 \pm 0.03$

Table 7.6: Fraction of SLT_e tags that are real electrons in the generic jet datasets.

We can see that there is little variation across η , and the variation of F_e^{tag} in isolation is of the same magnitude as the systematic uncertainty. The electron contamination does fluctuate high at low p_T , but this increase is only present in the jet20 dataset, and a different value for F_e^{tag} in this region is not warranted from K_s sample which we explore immediately following. Therefore, we make apply an overall correction to the fake matrices for the electron contamination. F_e^{tag} is 35%, 39%, and 45% for tag levels 1, 2, and 3, respectively. These corrections are also shown in Table 7.6.

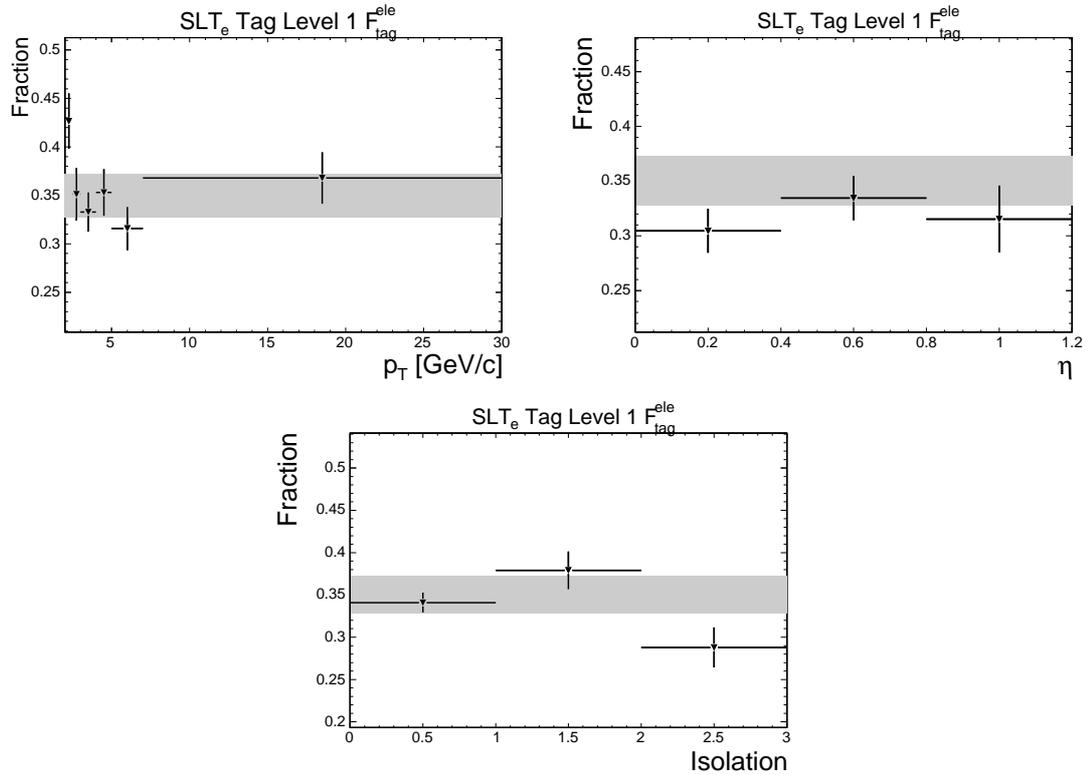


Figure 7.17: Weighted average of F_e^{tag} for tag level 1 combining generic jet datasets as a function of track p_T , η , and isolation bin in the fake matrix. Shown also is the average value with uncertainties.

7.3 Cross Checks and Systematics

In this section, we use the fake matrix, corrected by F_e^{tag} , to predict the number of tags in a sample of charged pions from K_s . We look at predictions and measurements of tags in this sample as a function of p_T , η , and isolation to validate our subtraction methodology.

We select pions from K_s decay in the generic jet datasets. We use the same criteria as the generic track selection, but we do not attempt to address the jet trigger bias and the E_{frac} requirement is released. K_s candidates are required to be matched

to a jet ($\Delta R \leq 0.4$). In addition to these requirements, we also require that the track has an oppositely-charged partner track which forms a vertex. The vertex is determined through a constrained fit of the track parameters. An invariant mass window of (0.4875, 0.5075) with a side-band of (0.55, 0.57) is selected. The track pair must have an L_{xy} significance greater than 8. Figure 7.18 shows the L_{xy} significance and invariant mass of the track pairs selected. We find 63623 events in the signal region and 18164 in the sidebands. The selection criteria are summarized in Table 7.7.

K_s Reconstruction Requirements
SLT _e taggable track (see Table 4.6)
Oppositely-charged partner track with: Valid Reconstructed Vertex $0.4875 < M_{invariant} \text{ (GeV}/c^2) < 0.5075$ $L_{xy} \text{ significance} > 8$

Table 7.7: Selection requirements for K_s in events triggered on a high E_T jet.

We apply the corrected fake matrices to the pions from K_s collected in this manner and compare the prediction against total number of tags measured. The agreement is well within the measurement errors for all three tag levels. Table 7.8 summaries the results. The predicted systematic uncertainty in the table includes the statistical uncertainty of the fake matrix bins, the overall 5% systematic uncertainty on the matrix, and the uncertainty in F_e^{tag} .

We can also compare the predicted and measured tags in the K_s sample as a function of p_T , η , isolation, and E_{frac} . Figure 7.19 show the (predicted-measured)/predicted as a function of these key variables for tag level 1 (the agreement for the other two

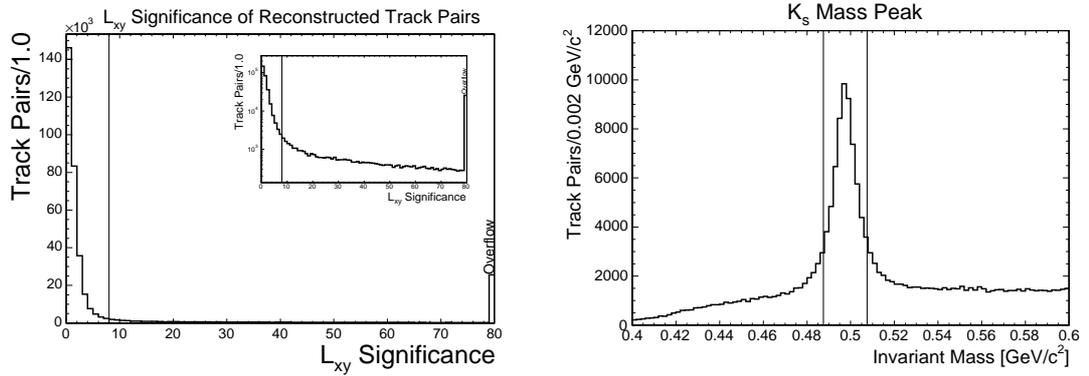


Figure 7.18: On the left is the L_{xy} significance of track pairs requiring $0.4875 < M_{invariant} (\text{GeV}/c^2) < 0.5075$. Embedded is the same plot with a logarithmic scale. On the right is the invariant mass of tracks pairs requiring L_{xy} significance > 8 . Track pairs are collected in the generic jet datasets.

Predicted versus Measured Tags in $K_s \rightarrow \pi^+\pi^-$ Sample			
Tag Level	Measured Tags	Predicted Tags	(Predicted-Measured)/Predicted
1	2556 ± 90 . stat	$2606. \pm 300$. syst	0.019 ± 0.035 stat ± 0.113 syst
2	1720 ± 78 . stat	$1820. \pm 210$. syst	0.055 ± 0.043 stat ± 0.109 syst
3	952 ± 61 . stat	$983. \pm 136$. syst	0.031 ± 0.062 stat ± 0.135 syst

Table 7.8: Predicted and Measured SLT_e tags in the K_s pion sample by tag level.

tag levels is similar). The agreement validates our application of the F_e^{tag} correction to all bins and verifies that the fake matrices can predict the behavior of fakes in a different sample.

7.4 Summary

In this chapter, we have made a fake matrix out tracks of the generic jet datasets. The matrix predicts the SLT_e tagging efficiency for generic tracks after calorimeter

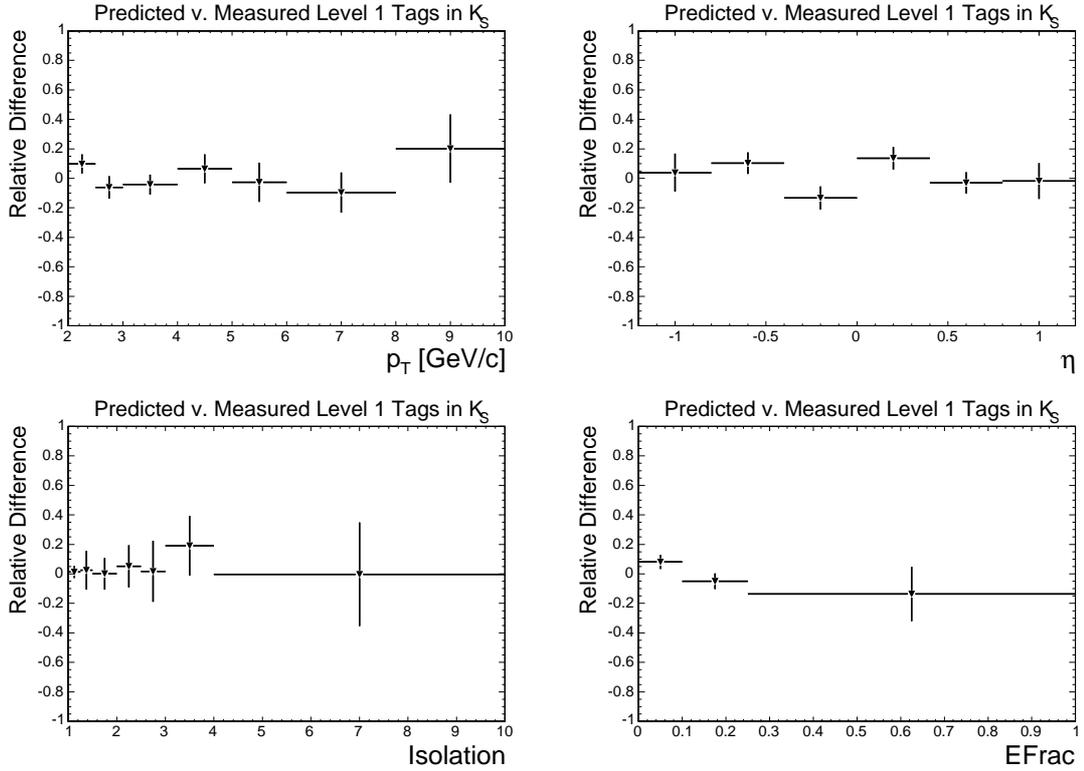


Figure 7.19: These need to be remade. $(\text{Predicted}-\text{Measured})/\text{Predicted}$ tags in a sample of pions from K_s as a function of p_T , η , isolation, and E_{frac} . SLT_e tag level 1 is used.

requirements have been imposed. We find that the fake tag rate in the generic jet datasets only agree with each other after requiring $E_{frac} < 0.25$. We estimate that the fake matrix may under-predict the efficiency in the $E_{frac} > 0.25$ region as much as $(21 \pm 21)\%$. We also find that the fake matrix correctly predicts the tagging rate for generic tracks in events triggered on a high E_T photon within 5%. After correcting for the electron contamination, we cross check our results with a sample of pions from K_s decay. We find that the tagging efficiency (tagged over taggable) for non-electrons is 5.8%, 4.1%, and 2.2%, for tag levels 1, 2, and 3, respectively in $t\bar{t}$ MC. The average

efficiency of calorimeter requirements is 8% in the same sample.

Chapter 8

SLT_e Cross Check in $b\bar{b}$ Events

In this chapter, we cross check the SLT_e tagger in $b\bar{b}$ events. In the previous chapters, we use a tag matrix and fake matrix to predict the tagging probability for electrons and fake electrons embedded in jets. We also measure the conversion efficiency and over-efficiency scale factors. Here, we combine all of these tools together, but the primary issue we will address is the reliability of the tag matrix to predict the SLT_e tagging rate in a non-isolated sample. In Chapter 6, we build the tag matrix using electrons from photon conversions. When we predict the tagging efficiency in a high p_T electron sample from Z bosons, we find that the tag matrix over-predicts the real tagging efficiency due to an enhancement of high E_T bremsstrahlung photons (see Table 6.3). We will find a similar effect here, where the density of photons and hadrons from the jet environment affects the predicted tag rate.

8.1 $b\bar{b}$ Sample Selection

We identify b jets by multiply tagging di-jet events with SECVTX as well as identifying a lepton with more than 8 GeV of transverse energy. The event selection begins by collecting events that fire the inclusive 8 GeV electron and muon triggers. The electron selection criteria are summarized in Table 8.1. This is the same criteria as in Table 4.1, but we additionally reject conversions, require the track to be close to the primary vertex in z , and require the electron to be non-isolated. The conversion filter and isolation variable used are described in Section 3.2.1. In this case, the isolation variable is a calorimeter-based isolation variable, not the typical isolation variable used for the SLT_e. The muon selection criteria are summarized in Table 8.2.

Trigger Electron Selection for $b\bar{b}$ Sample Construction
Fiducial to CES
$ q\Delta X < 3.0$ cm
$ \Delta Z < 5.0$ cm
$0.5 < E_{EM}/P < 2.0$
$E_{Had}/E_{EM} < 0.05$
$L_{shr} < 0.2$
$\chi_{strip}^2 < 10$
$E_T > 9.0$ GeV
$p_T > 8.0$ GeV/ c
Not a conversion
$Isol > 0.1$
$ z_0 - z_{vtx} < 5.0$ cm

Table 8.1: 8 GeV trigger electron selection criteria for the $b\bar{b}$ sample construction.

After the trigger lepton is identified, we require the lepton to be matched to a jet ($\Delta R \leq 0.4$) that has a loose SECVTX tag. We call this jet the *lepton jet*. An *away*

 Trigger Muon Selection for $b\bar{b}$ Sample Construction

Fiducial to CMU and CMP

$$\begin{aligned}
 p_T &> 8.0 \text{ GeV}/c \\
 |\Delta x|_{CMU} &< 10.0 \text{ cm} \\
 |\Delta x|_{CMP} &< 15.0 \text{ cm} \\
 Isol &> 0.1 \\
 |z_0 - z_{vtx}| &< 5.0 \text{ cm}
 \end{aligned}$$

Table 8.2: Muon selection criteria. The $Isol$ variable is the standard calorimeter isolation, not the track-based isolation normally used for the SLT_e .

jet is required to be opposite the lepton jet, with $|\Delta\phi| > 2.0$ and $|\Delta\eta_D| < 1.5$. The away jet must also have a tight SECVTX tag. If more than one jet meets these requirements, we use the one with the largest $|\Delta\phi|$. Again, all jets are clustered with the JetClu algorithm described in 3.3.1, have corrected $E_T > 20$ GeV and $|\eta_D| < 2.0$. This event selection is illustrated in Figure 8.1.

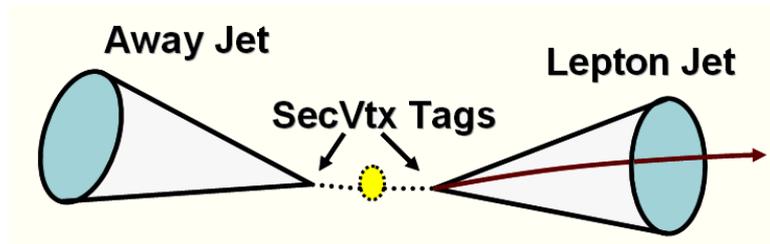


Figure 8.1: Illustration of the $b\bar{b}$ event selection. We look for a Loose SECVTX tag and an 8 GeV lepton in the lepton jet. We look for a Tight SECVTX tag in the away jet. The SLT_e tagging efficiency is measured in the away jet.

We find a total of 33255/68616 away jets in the electron/muon trigger samples. We use the identical event selection in an 8 GeV electron/muon MC sample. This MC sample is a Pythia di-jet sample filtered for an 8 GeV electron or muon at generator level. We find 6532/8888 away jets in the electron/muon MC. Since we have nearly

seven times more events in data than MC, we keep track of the statistical uncertainty in the MC, as well.

We measure the efficiency to tag the away jet in both data and MC. We assume that the presence of a tight $SECVTX$ tag does not affect the SLT_e tag rate differently in data and MC. We also assume that our sample selection results in relatively pure, comparable samples of $b\bar{b}$ in data and MC. In MC, we find that approximately 2% of the events are light jets, and 3% are c -jets (the rest are b -jets) after we have the event selection described. A comparison of the beamline corrected d'_0 distribution of taggable SLT_e tracks in the away jet between data and MC in Figure 8.2 shows good agreement. Conversions have been rejected with the SLT_e conversion filter. The agreement with MC suggests that data is comparably b -enriched.

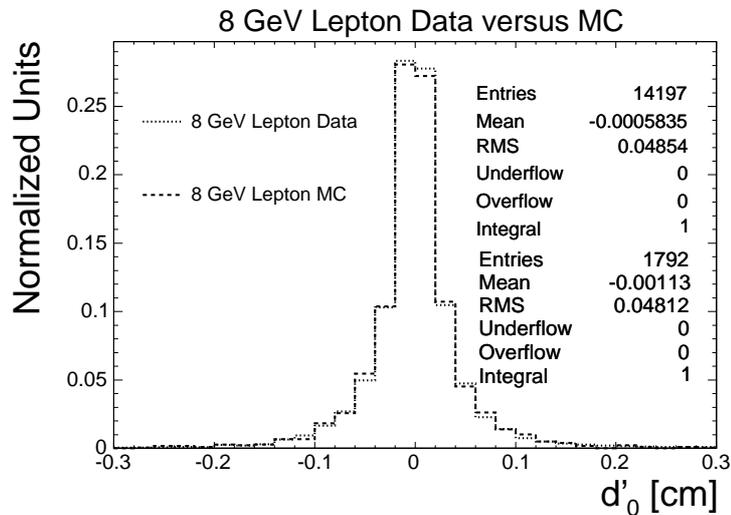


Figure 8.2: Comparison of the beamline corrected impact parameter of taggable SLT_e tracks in the away jet of the $b\bar{b}$ sample. $b\bar{b}$ events are collected in the 8 GeV lepton samples.

8.2 Application of the SLT_e Tagger

We apply the SLT_e tagger to the tracks in the away jet. If at least one track is tagged in the jet, then we consider the jet to be tagged. We consider a track taggable/tagged in data if it passes the requirements listed in Table 4.6/4.7. In MC, we use the tag matrix, fake matrix, and conversion efficiency/over-efficiency scale factors to determine the probability to tag a track given its species.

The species of a track is determined by matching the reconstructed track hits to the hits simulated by the MC. Failing a proper match, the track's helix parameters are matched to the those of the simulated particle's true parameters. Failing this match, the simulated particle closest to the track in ΔR space with the correct charge is chosen as the true particle species. This last step is only performed for $\sim 1\%$ of all tracks.

In the MC, the probability to tag a track is determined from the tag matrix if the particle is an electron, or the fake matrix otherwise. The value from the fake matrix is corrected for the electron contamination of the generic jet samples, F_{ele}^{tag} . If the particle is identified as a conversion then the probability to tag is re-weighted by $1 - SF$, where SF is the conversion efficiency or over-efficiency scale factor. The SF used is determined by whether or not the particle originated from a conversion photon or not. The probability to tag a jet is

$$P_{jet} = 1 - \prod_i (1 - p_i) \quad (8.1)$$

where p_i is the probability to tag a track.

8.3 Data/MC Comparison

Table 8.3 shows the results from applying the SLT_e tagger to the away jets. The efficiency quoted is the number of tags per jet, which includes the acceptance and branching fraction. The first uncertainty is the statistical and second is the systematic. We have broken down the results by dataset as well as by SLT_e tag level and shown the absolute differences between data and MC. For MC, we show the contributions from HF electrons, fake electrons, and conversion electrons. Other electrons (neither from HF nor conversions) are nearly negligible and are bundled with the fakes.

Sample	Total ε (%)	HF Electron ε (%)	Conversion ε (%)	Fake + Other ε (%)
SLT_e Tag Level 1				
Ele. Data	$4.47 \pm 0.11 \pm 0.00$	—	—	—
Ele. MC	$4.37 \pm 0.12 \pm 0.18$	$3.04 \pm 0.18 \pm 0.11$	$0.32 \pm 0.03 \pm 0.02$	$1.05 \pm 0.03 \pm 0.07$
Ele. Δ	$0.10 \pm 0.17 \pm 0.18$	—	—	—
Muo. Data	$4.17 \pm 0.08 \pm 0.00$	—	—	—
Muo. MC	$4.36 \pm 0.10 \pm 0.18$	$3.08 \pm 0.15 \pm 0.11$	$0.30 \pm 0.03 \pm 0.02$	$1.01 \pm 0.03 \pm 0.07$
Muo. Δ	$-0.19 \pm 0.13 \pm 0.18$	—	—	—
Weighted Δ	$-0.10 \pm 0.10 \pm 0.18$	—	—	—
SLT_e Tag Level 2				
Ele. Data	$3.61 \pm 0.10 \pm 0.00$	—	—	—
Ele. MC	$3.64 \pm 0.10 \pm 0.15$	$2.65 \pm 0.15 \pm 0.10$	$0.28 \pm 0.03 \pm 0.02$	$0.74 \pm 0.02 \pm 0.05$
Ele. Δ	$-0.03 \pm 0.14 \pm 0.15$	—	—	—
Muo. Data	$3.37 \pm 0.07 \pm 0.00$	—	—	—
Muo. MC	$3.62 \pm 0.09 \pm 0.15$	$2.66 \pm 0.13 \pm 0.10$	$0.26 \pm 0.02 \pm 0.02$	$0.72 \pm 0.02 \pm 0.05$
Muo. Δ	$-0.25 \pm 0.10 \pm 0.15$	—	—	—
Weighted Δ	$-0.18 \pm 0.08 \pm 0.15$	—	—	—
SLT_e Tag Level 3				
Ele. Data	$2.43 \pm 0.08 \pm 0.00$	—	—	—
Ele. MC	$2.58 \pm 0.07 \pm 0.11$	$1.98 \pm 0.12 \pm 0.08$	$0.21 \pm 0.02 \pm 0.01$	$0.40 \pm 0.01 \pm 0.03$
Ele. Δ	$-0.15 \pm 0.11 \pm 0.11$	—	—	—
Muo. Data	$2.26 \pm 0.06 \pm 0.00$	—	—	—
Muo. MC	$2.55 \pm 0.06 \pm 0.11$	$1.97 \pm 0.10 \pm 0.07$	$0.20 \pm 0.02 \pm 0.01$	$0.39 \pm 0.01 \pm 0.03$
Muo. Δ	$-0.29 \pm 0.08 \pm 0.11$	—	—	—
Weighted Δ	$-0.24 \pm 0.07 \pm 0.11$	—	—	—

Table 8.3: Per jet tagging rate in data and MC, for the inclusive electron and muon datasets, and by SLT_e tag level.

We consider several sources of systematic uncertainties in the application of the tagger in MC. We list them in approximately decreasing order of importance to the b -tagging efficiency measured in this note. Since the values depend on the tag level and the sample it is measured, we give only the approximate contributions to the systematic uncertainty here. The total systematic uncertainty is shown in Table 8.3.

- ($\sim 2.5\%$). Uncertainty on the tag matrix prediction. This is dominated by the 3% systematic applied to cover the prediction variation in a number of key variables. Note that only 70 – 80% of the tracks are electrons for which this systematic is applied. Other contributions include the statistical uncertainty per bin and the high p_T correction systematics (see Chapter 6).
- ($\sim 2.5\%$). Uncertainty on the calorimeter variables (see Section 6.4).
- ($\sim 2.0\%$). Uncertainty on the fake matrix prediction. This is due to the systematic on the matrix prediction as well as the estimate of the electron contamination (see Chapter 7).
- ($\sim 1.8\%$). Conversion filter over-efficiency SF systematic (see Section 5.4).
- ($\sim 0.5\%$). Conversion filter efficiency SF systematic (see Section 5.5).

As can be seen from Table 8.3, the efficiency is over-predicted in MC, and the over-prediction increases with tag level. The reason for this over-prediction is because the local environment of the electrons in the HF sample is not the same as the electrons in the conversion sample we used to build the tag matrix. Although we parameterize the tag matrix by isolation, the definition of isolation cannot accommodate neutral

particles which can have a large impact on the tag rate. We have seen this same effect before with electrons from Z bosons in Chapter 6. The presence of a photon from bremsstrahlung was not accounted for by the isolation variable and distorted the CES electron ID quantities. In addition, the tighter the requirement is on the likelihood, the more sensitive the measured efficiency was to small variations from the presence of unaccounted neutral particles.

Associating the entire difference between data and MC with a discrepancy in the tag matrix, we find that the electron contribution is over-predicted 3.0%, 6.2%, and 11.2%, for tag levels 1, 2, and 3, respectively. These numbers can be calculated by determining how much the total HF electron and conversion electron contribution must be scaled in order for the average predicted and measured values to agree. Therefore, we make a multiplicative correction to the tag matrix of 0.985 ± 0.029 , 0.969 ± 0.031 , and 0.944 ± 0.056 for each where the systematics have been inflated to cover the full variation. These numbers are one minus half the over-prediction. Table 8.4 shows the new tag rates with the corrected tag matrix.

Tag Level	Data ε (%)	MC ε (%) (corr.)
Inclusive Electron Dataset		
1	4.47 ± 0.11	$4.32 \pm 0.12 \pm 0.18$
2	3.61 ± 0.10	$3.54 \pm 0.10 \pm 0.17$
3	2.43 ± 0.08	$2.46 \pm 0.07 \pm 0.16$
Inclusive Muon Dataset		
1	4.17 ± 0.08	$4.31 \pm 0.10 \pm 0.18$
2	3.37 ± 0.07	$3.52 \pm 0.08 \pm 0.17$
3	2.26 ± 0.06	$2.43 \pm 0.06 \pm 0.16$

Table 8.4: Away jet tagging rate in data and MC with the new tag matrix correction applied.

Finally, we present in Figures 8.3 through 8.20 the distributions of away jet tags in data and MC after the tag matrix has been corrected. The predicted distributions are calculated in the same way as Table 8.3, using the (now corrected) Tag Matrix, Fake Matrix, and conversion SFs to determine the tagging probability of a track. Both electron and muon distributions are shown. Also shown are the distributions broken down between events where the SLT_e tag and the trigger lepton have either the same or opposite charge. Plots with jet quantities are per track not per jet, that is, one jet may have more than one entry. The last bin includes the overflow. Statistical uncertainties on the MC are shown. The agreement is quite strong in most distributions. This measurement gives us confidence that the SLT_e tagger will be well-behaved in the lepton+jets sample.

The isolation variable in Figures 8.4, 8.10, 8.16, is the same SLT_e isolation variable defined in Equation 4.1. The invariant mass shown in Figures 8.6, 8.12, and 8.18, is calculated between the SLT_e tag and the 8 GeV lepton (either electron or muon). The agreement in opposite-sign and same-sign distributions as well as both 8 GeV electron and muon distributions suggests little contamination from non- $b\bar{b}$ low p_T dilepton events such as Drell-Yan. The Jet E_T spectrum is well predicted in the 8 GeV electron dataset but not as well in the muon dataset, particularly for low E_T jets. This may be related to the fact that muons minimum ionize in the calorimeter lowering the reconstructed jet energy. The agreement at high E_T suggests that the environmental correction for the electron does not have a strong E_T dependence.

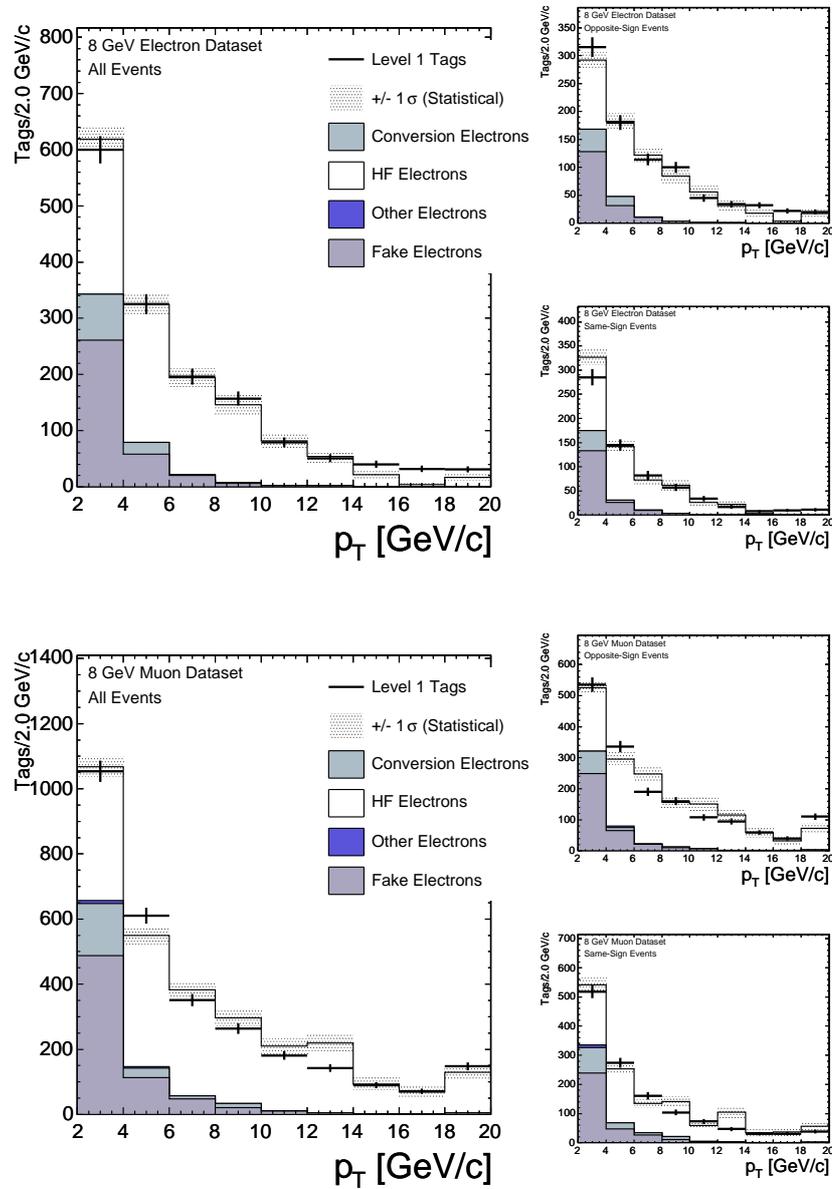


Figure 8.3: Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the track p_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

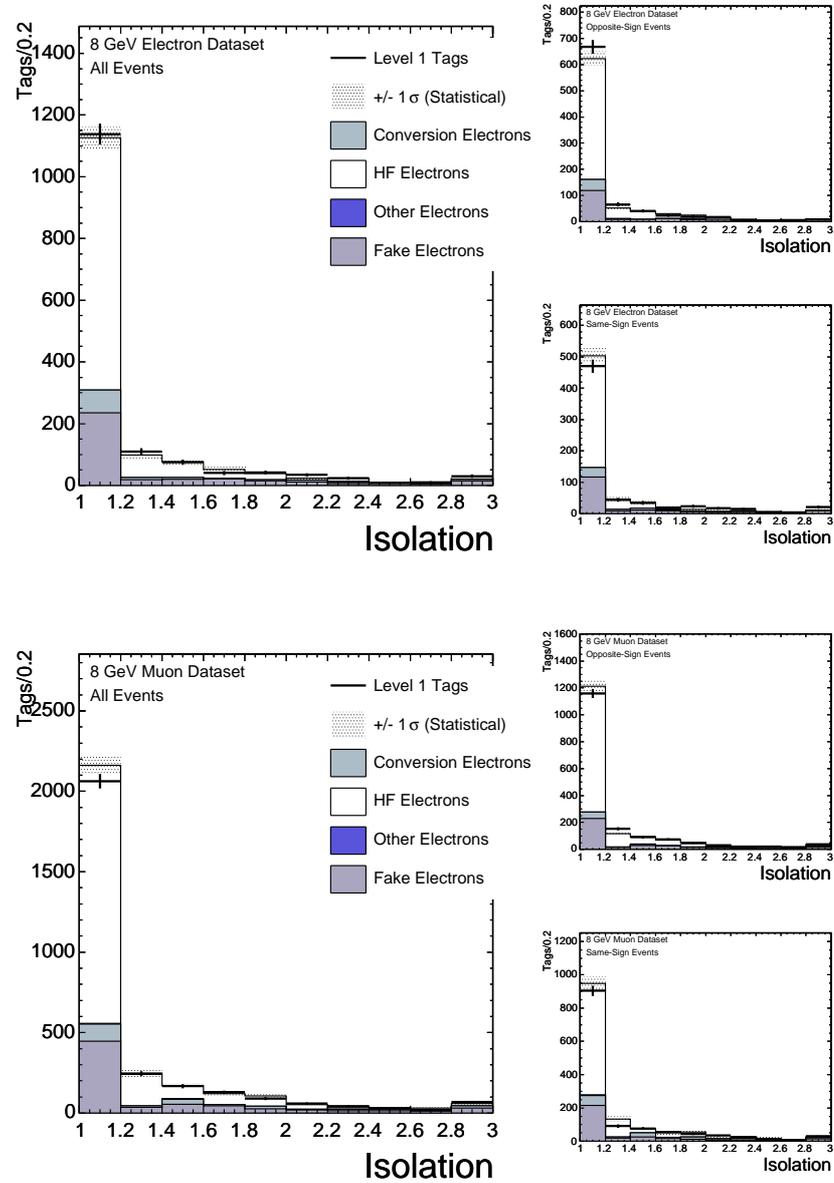


Figure 8.4: Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the isolation. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

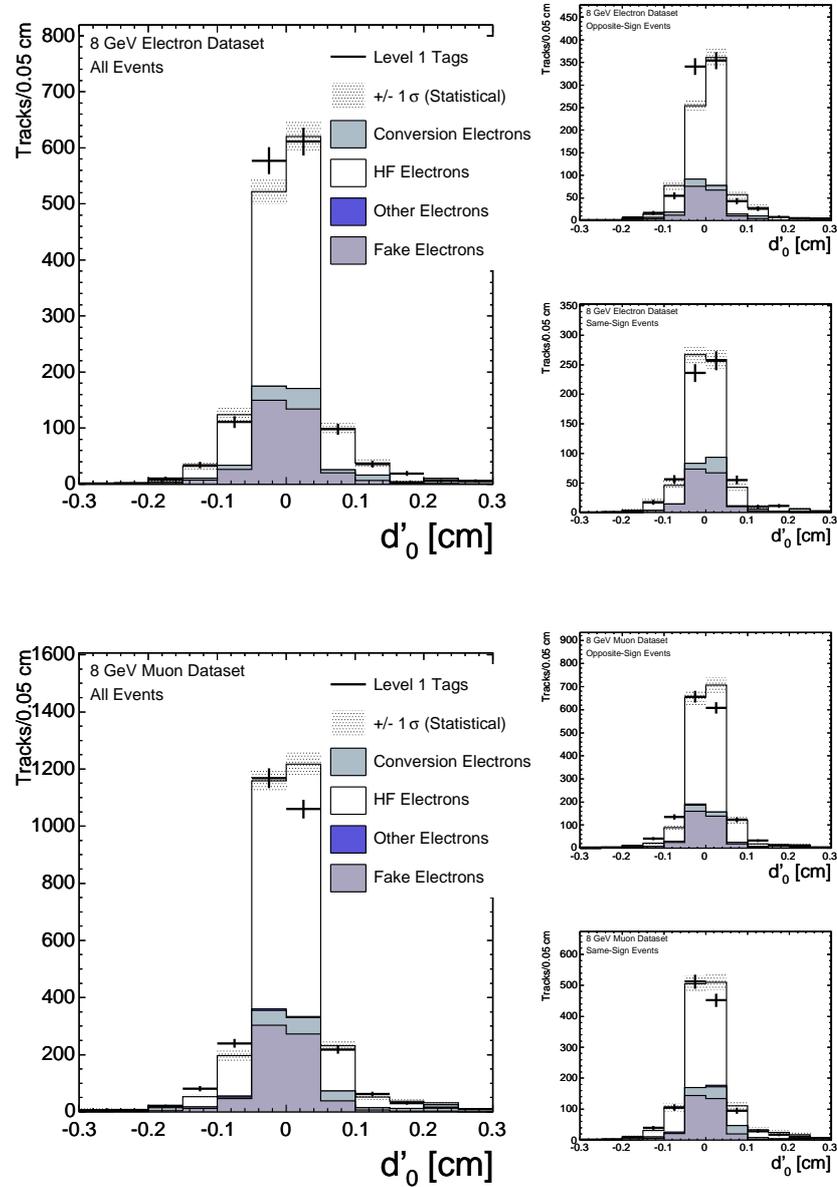


Figure 8.5: Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the beamline corrected impact parameter. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

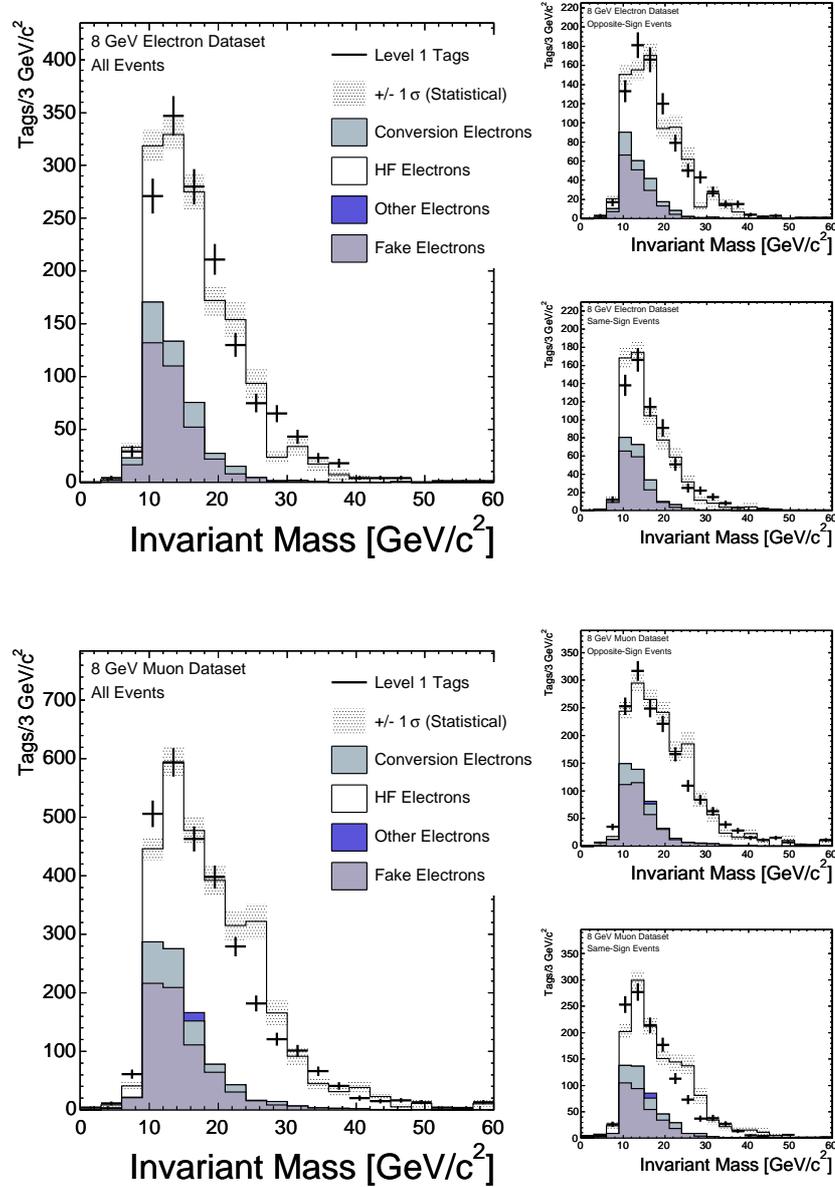


Figure 8.6: Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the invariant mass between the trigger lepton and the SLT_e tag. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

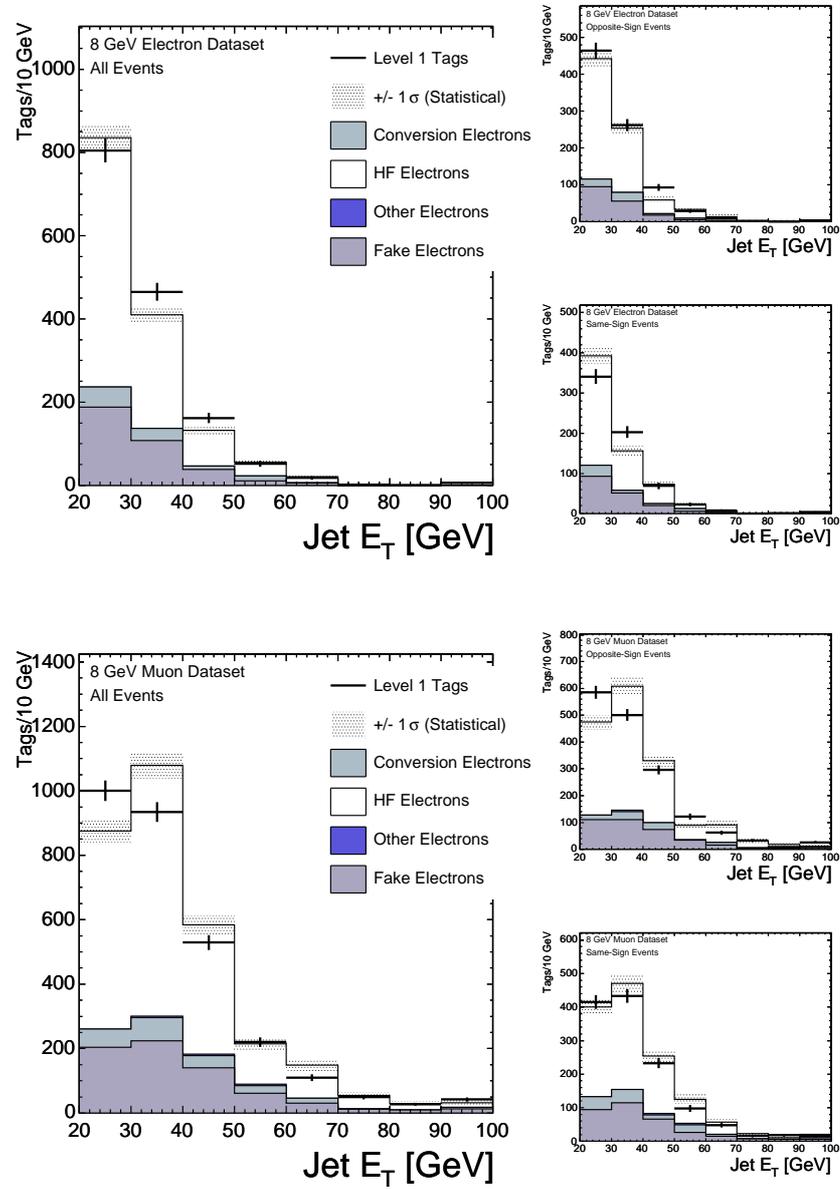


Figure 8.7: Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the jet E_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

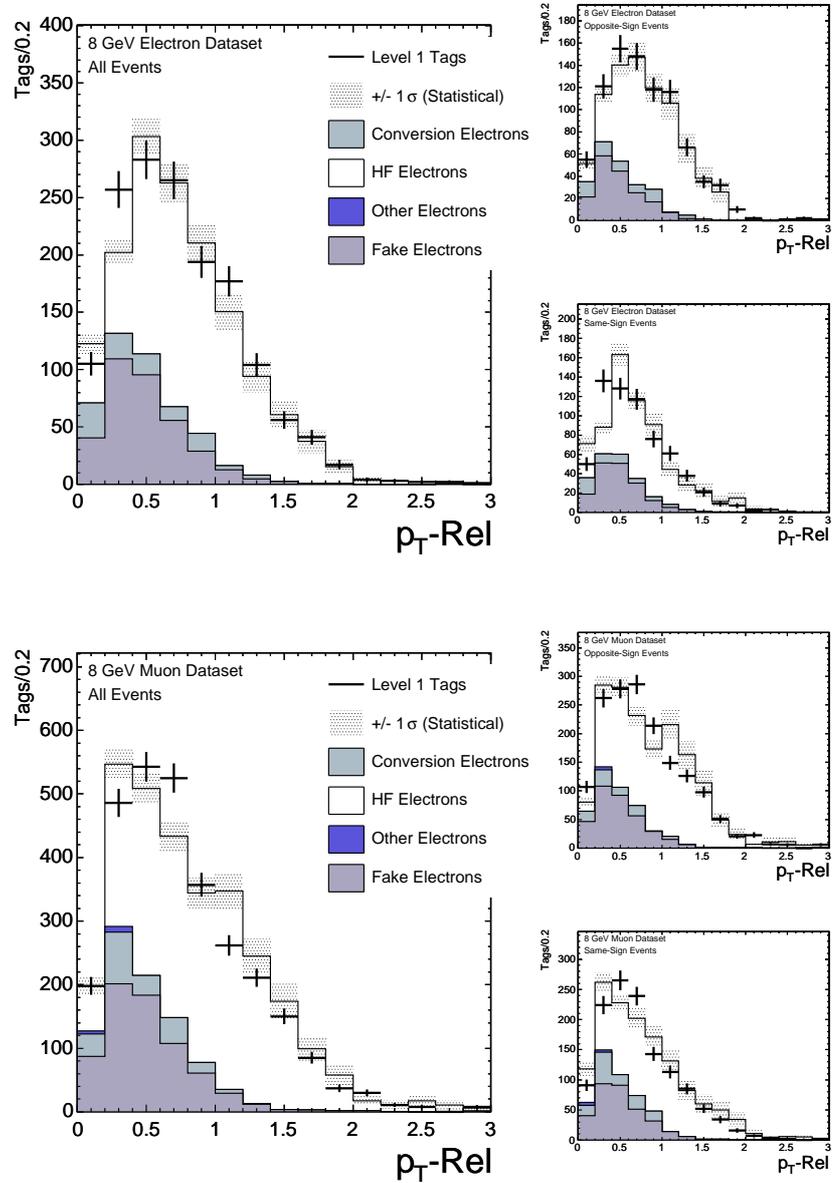


Figure 8.8: Measured and predicted SLT_e level 1 tags in a $b\bar{b}$ sample as a function of the relative p_T between the track and the jet. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

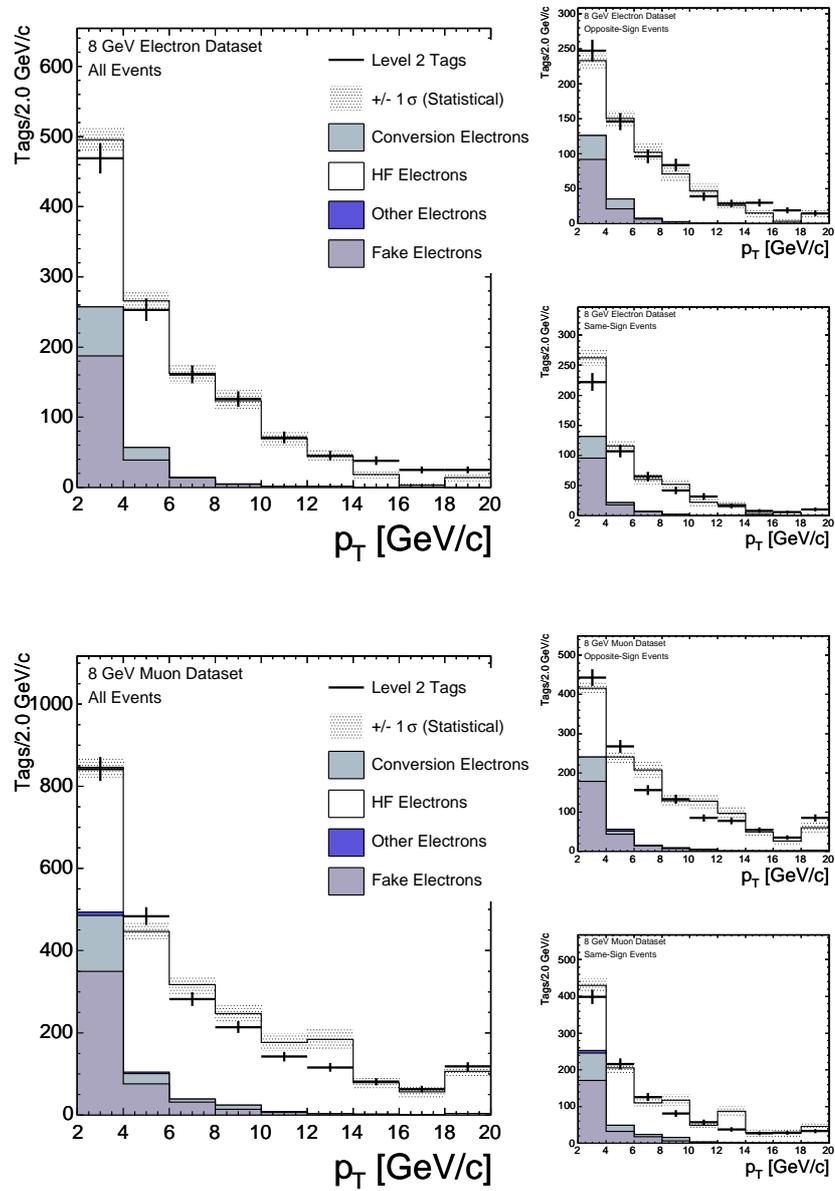


Figure 8.9: Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the track p_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

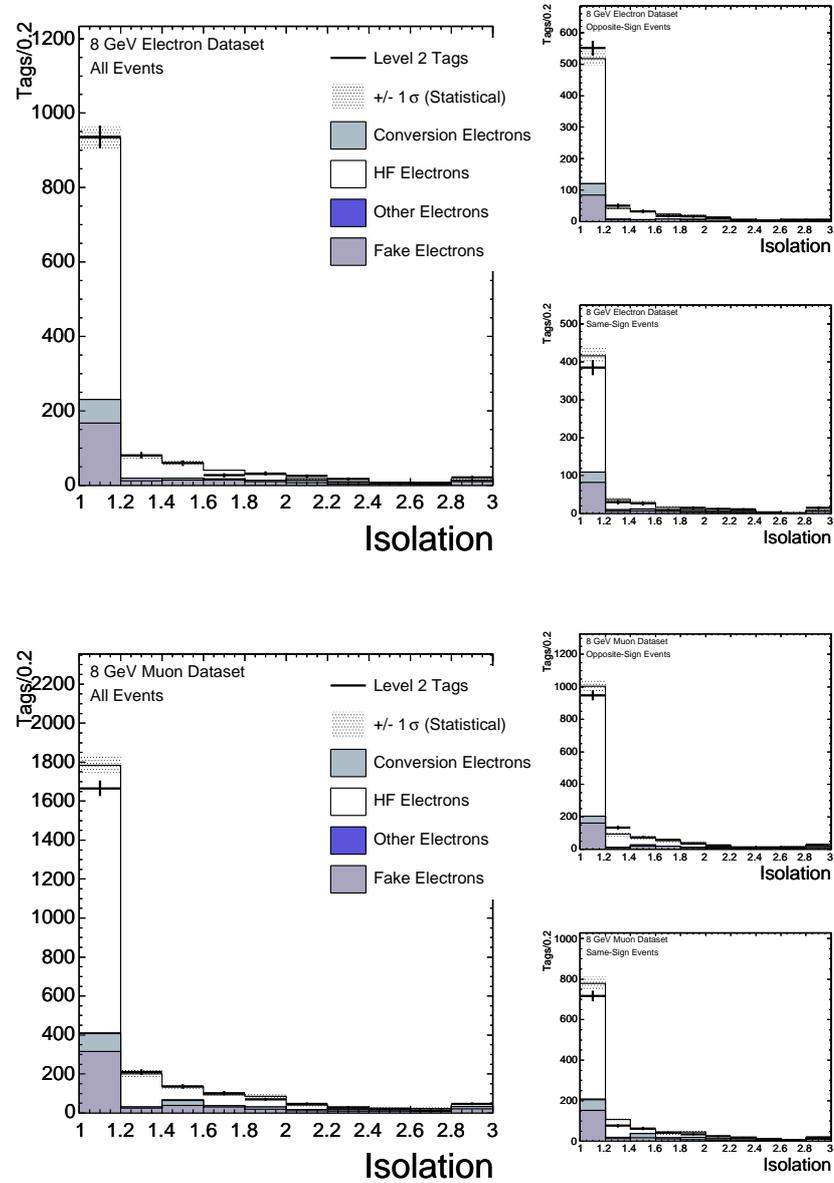


Figure 8.10: Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the isolation. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

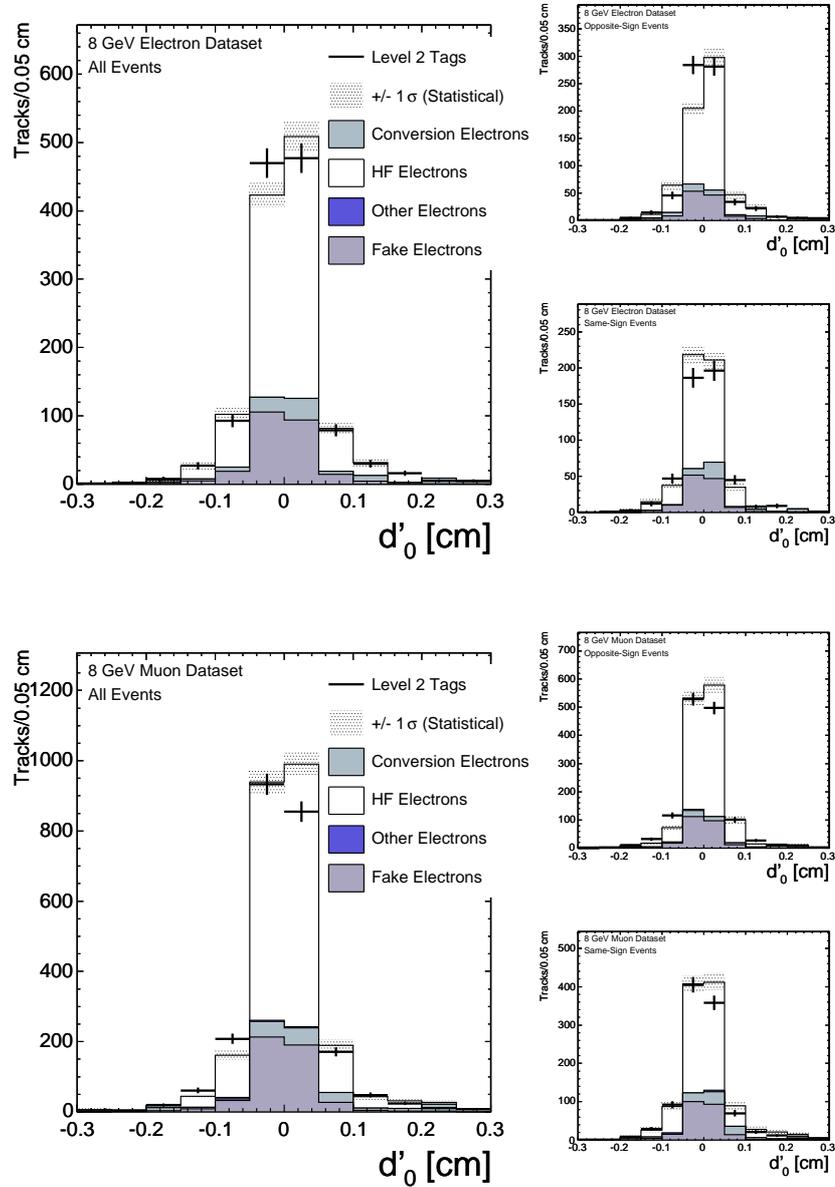


Figure 8.11: Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the beamline corrected impact parameter. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

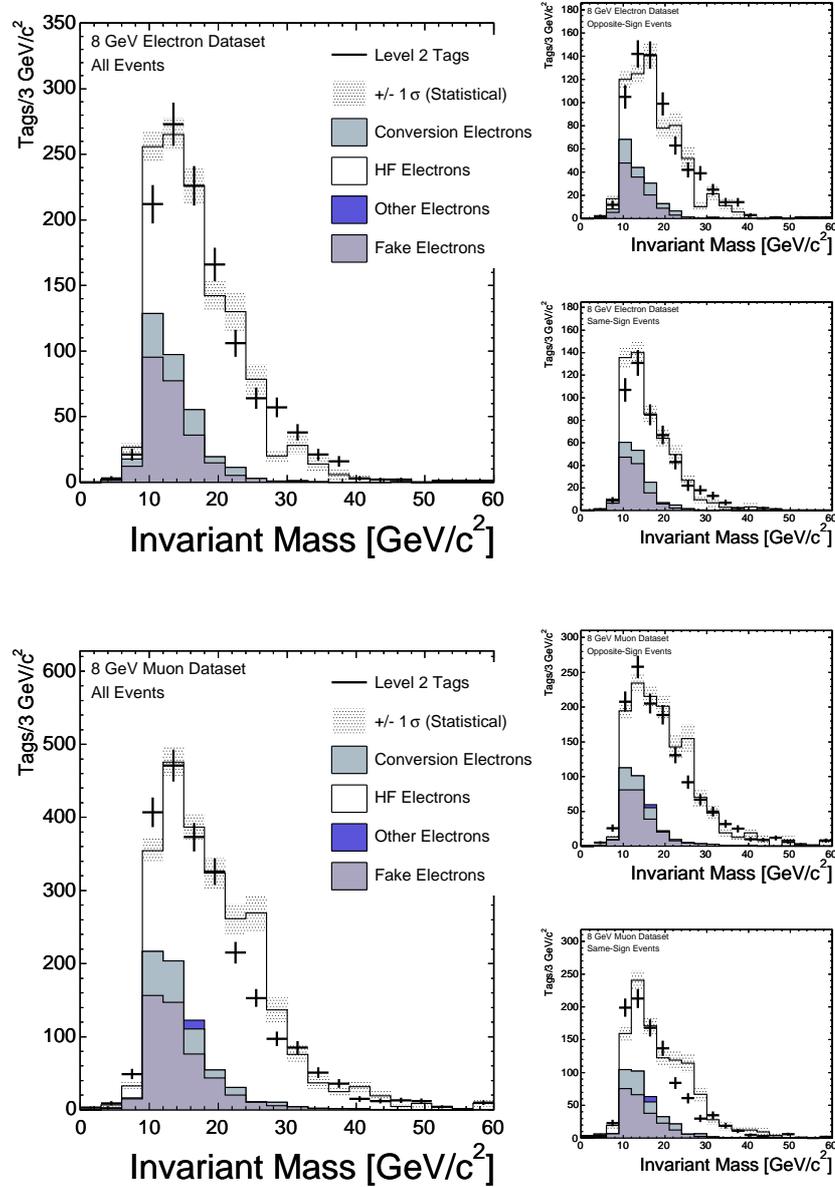


Figure 8.12: Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the invariant mass between the trigger lepton and the SLT_e tag. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

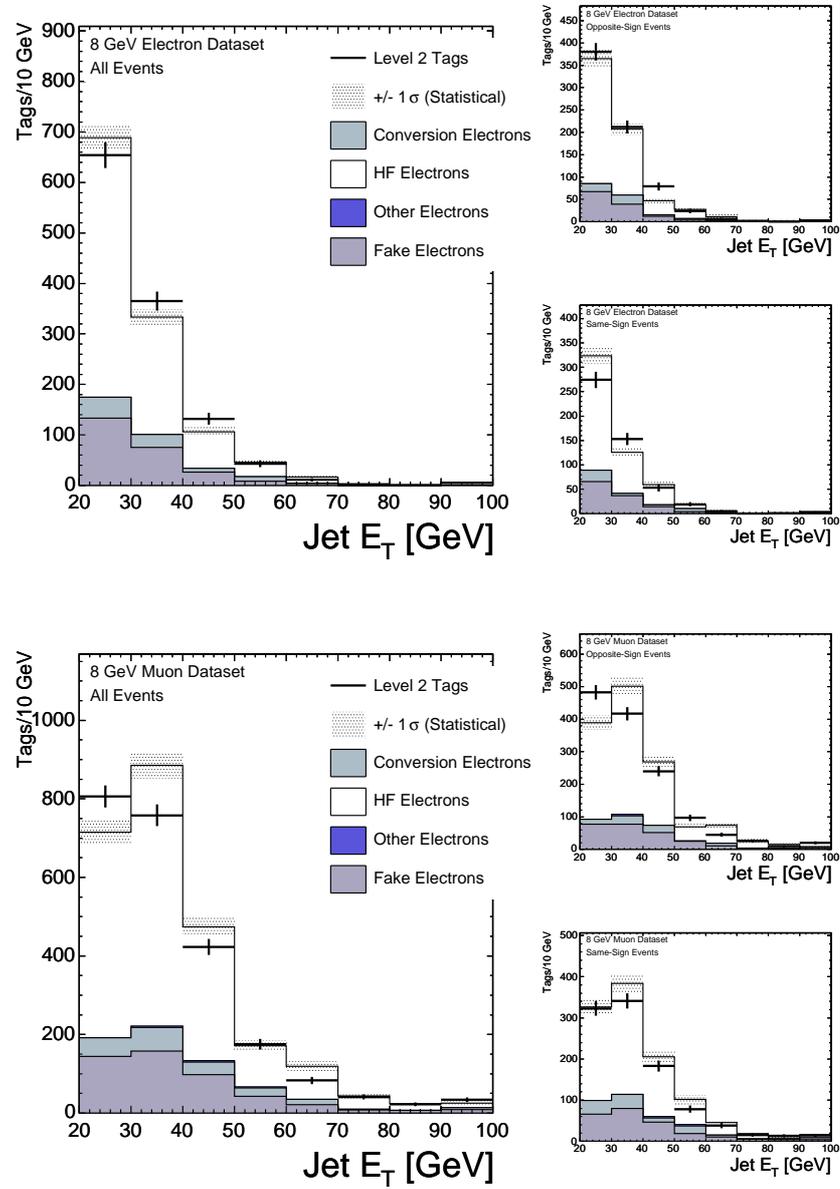


Figure 8.13: Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the jet E_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

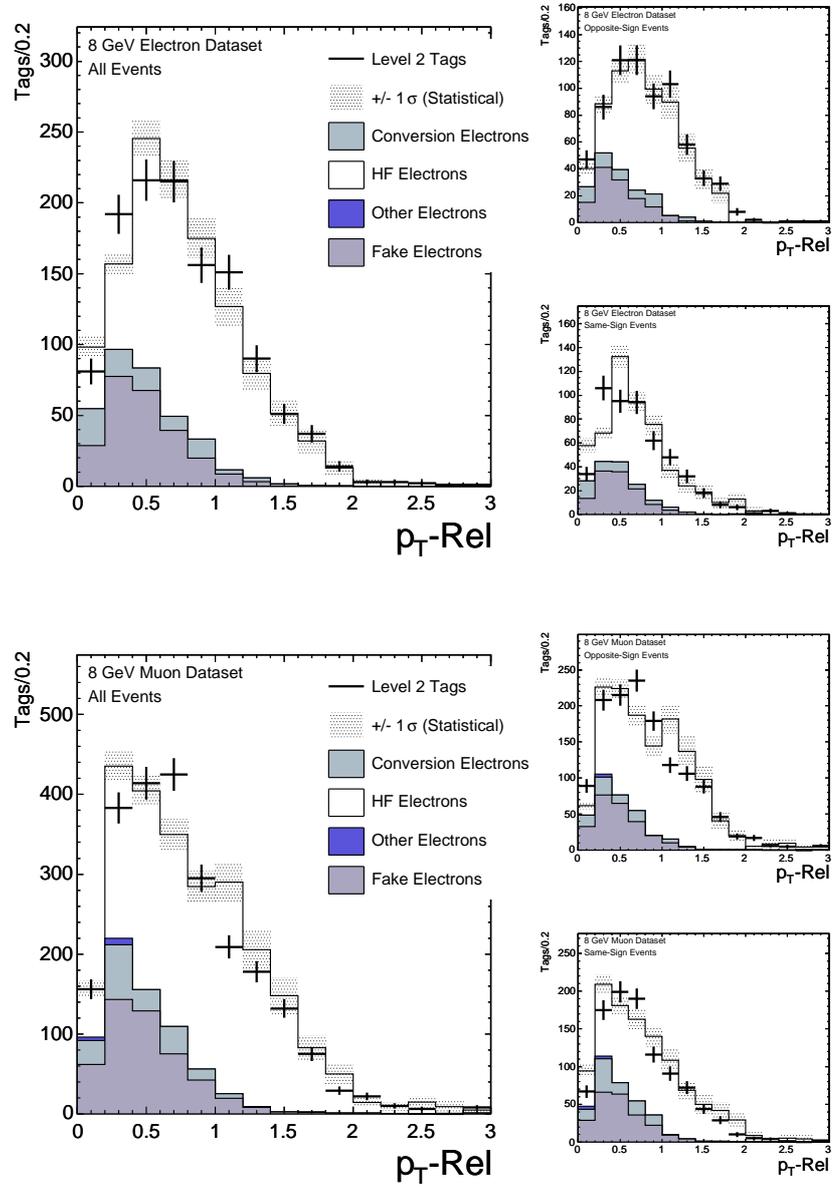


Figure 8.14: Measured and predicted SLT_e level 2 tags in a $b\bar{b}$ sample as a function of the relative p_T between the track and the jet. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

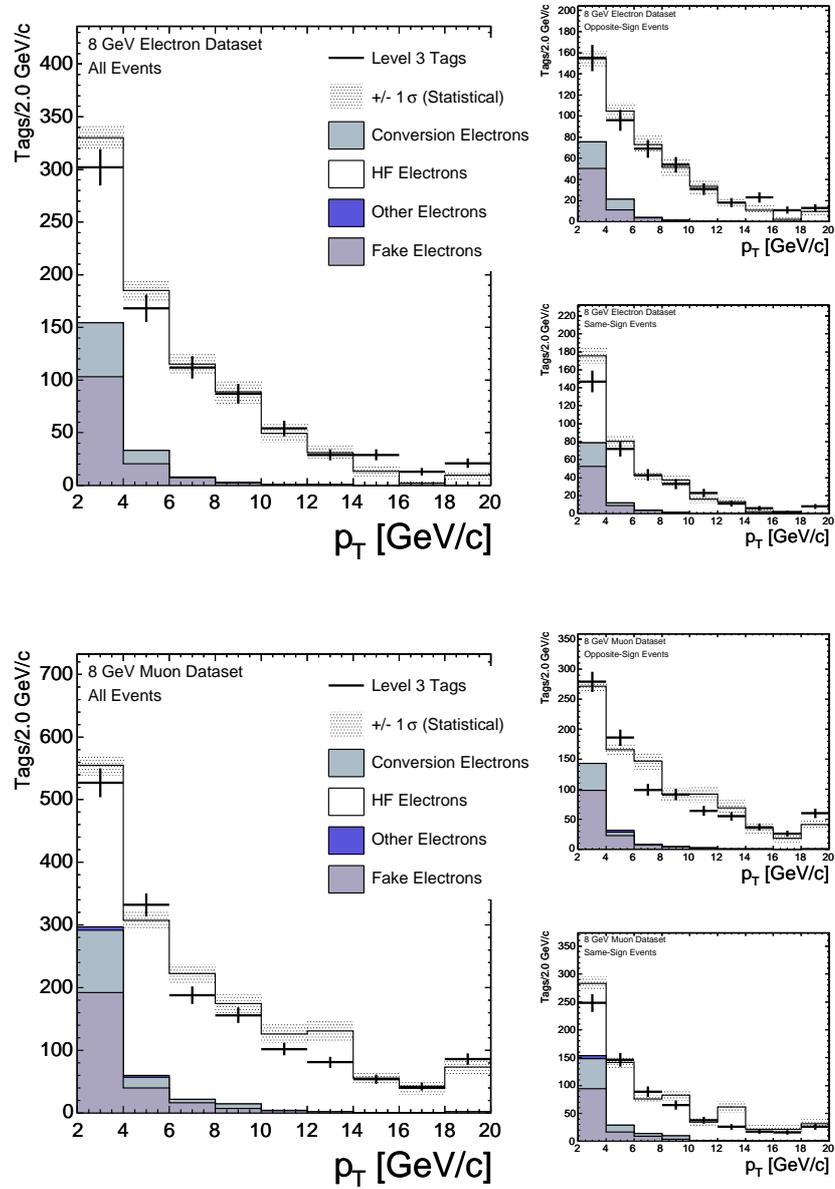


Figure 8.15: Measured and predicted SLT_e level 3 tags in a $b\bar{b}$ sample as a function of the track p_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

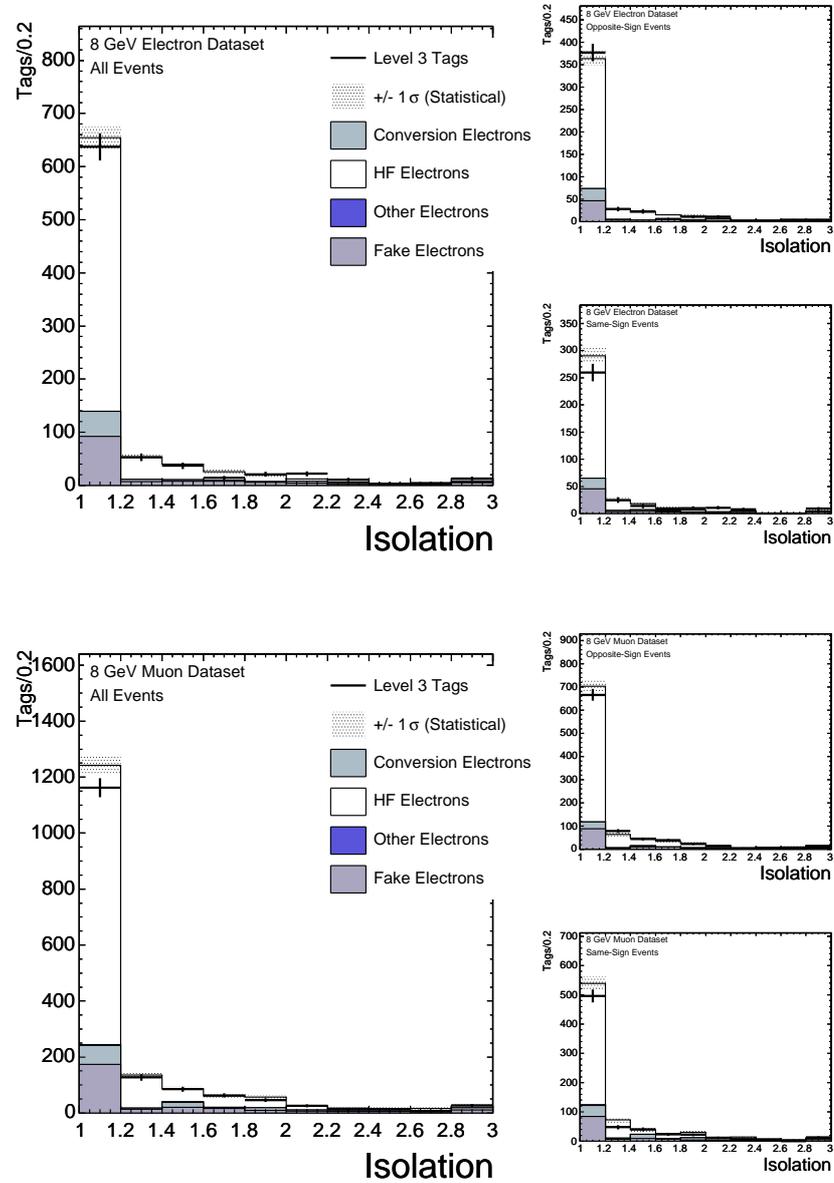


Figure 8.16: Measured and predicted SLT_e level 3 tags in a $b\bar{b}$ sample as a function of the isolation. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

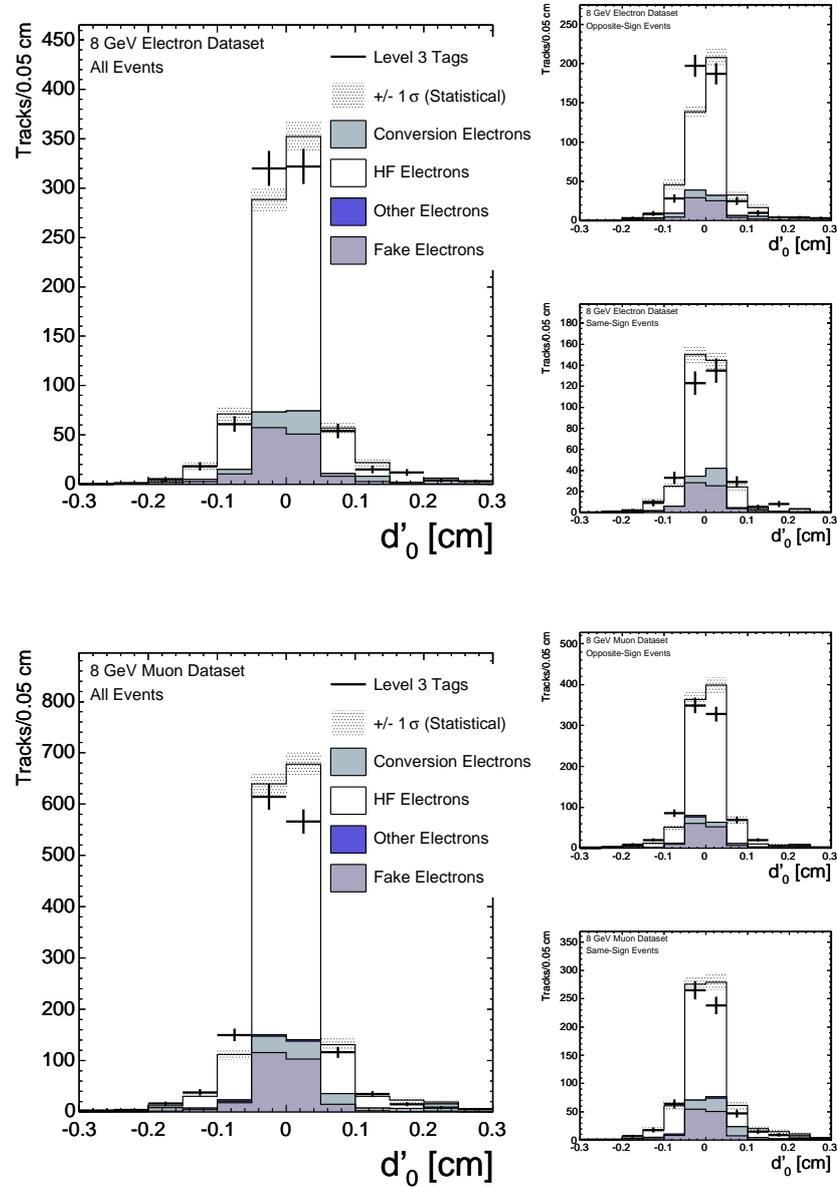


Figure 8.17: Measured and predicted SLT_e level 3 tags in a $b\bar{b}$ sample as a function of the beamline corrected impact parameter. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

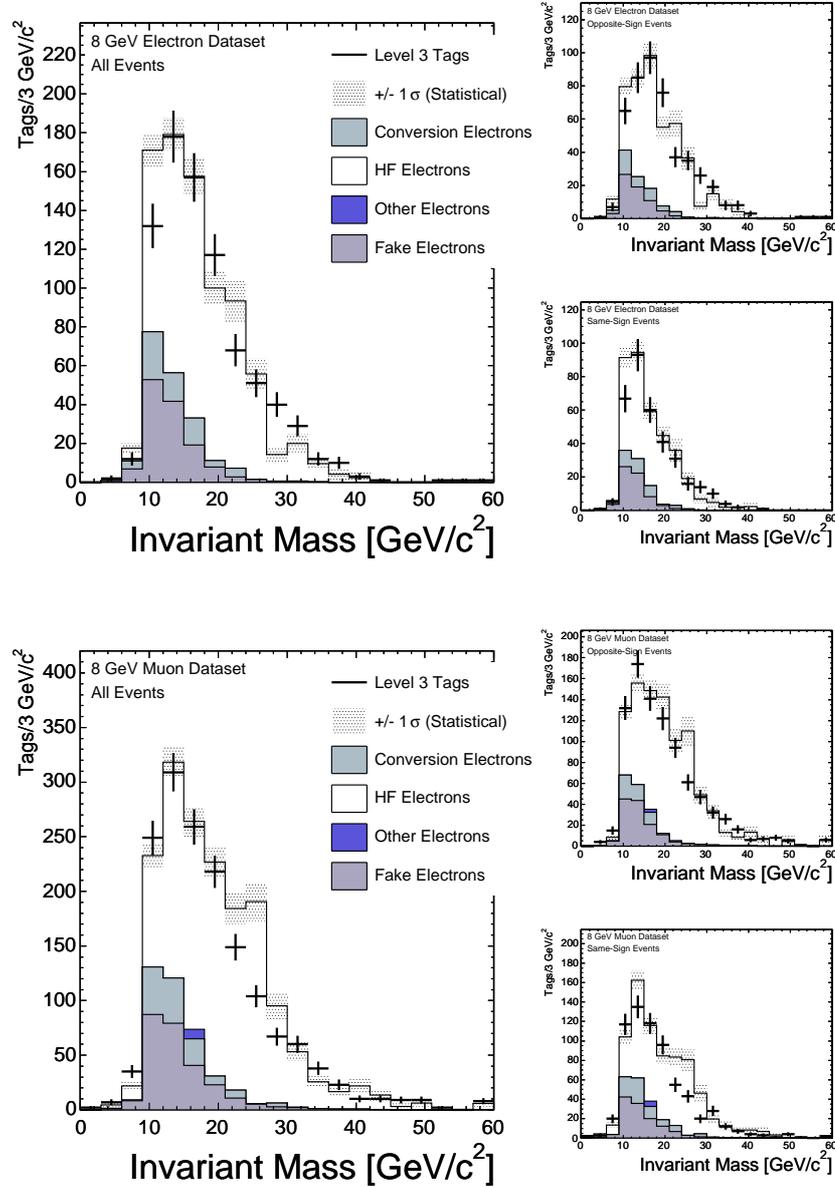


Figure 8.18: Measured and predicted SLT_e level 3 tags in a $b\bar{b}$ sample as a function of the invariant mass between the trigger lepton and the SLT_e tag. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

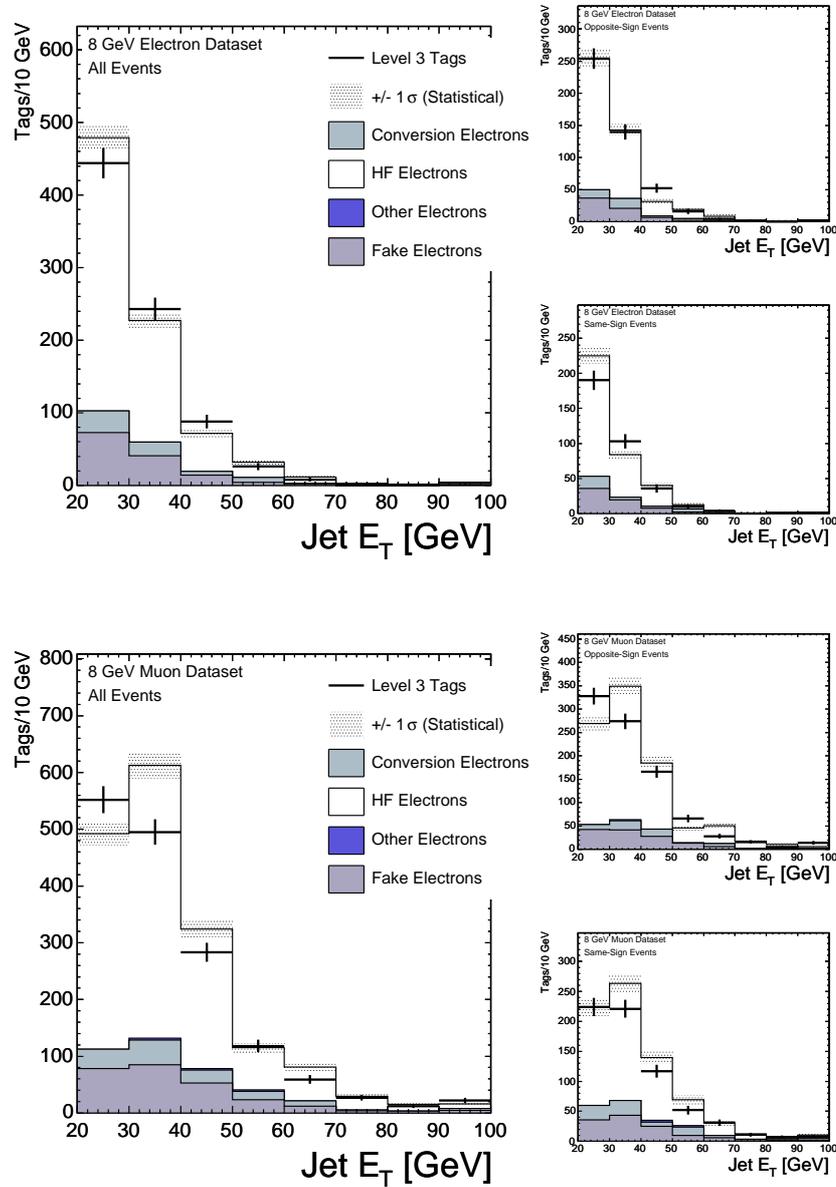


Figure 8.19: Measured and predicted SLT_e level 3 tags in a $b\bar{b}$ sample as a function of the jet E_T . Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

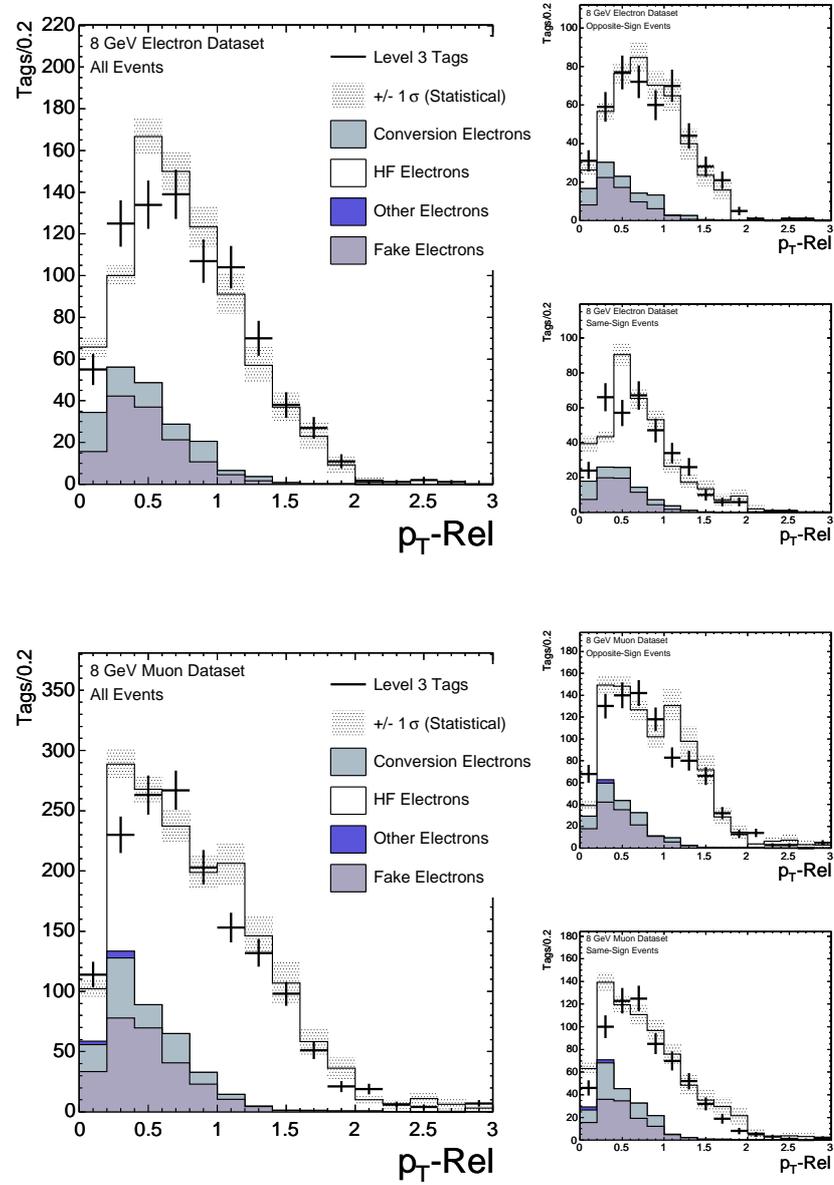


Figure 8.20: Measured and predicted SLT_e level 3 tags in a $b\bar{b}$ sample as a function of the relative p_T between the track and the jet. Events from the inclusive 8 GeV electron dataset are shown on top and those from the inclusive 8 GeV muon dataset are shown on the bottom.

Chapter 9

Cross Section

In this chapter, we present the measurement of the $t\bar{t}$ cross section using soft electron tags. We measure the cross section in the lepton+jets channel, described in Section 1.2.4, using 1.7 fb^{-1} of data. The primary background to $t\bar{t}$ events in this channel is W bosons produced in association with jets. Rather than rely completely on MC, the overall normalization of this background is calibrated with data, a procedure known at CDF as “Method 2.”

9.1 Event Selection

Events are selected through high E_T/p_T central electron/muon triggers. The offline selection criteria for the electron and muon are described in Section 3.2. Electrons/muons are required to have E_T/p_T greater than $20 \text{ GeV GeV}/c$ and be isolated. We also reject cosmic ray muons, conversion electrons, and Z bosons. Only one high p_T lepton can be reconstructed in the lepton+jets sample, and the flavor of that lep-

ton must be consistent with the trigger path. Jets are required to have a corrected¹ $E_T > 20$ GeV and have $|\eta_D| < 2.0$, where η_D is the detector η . We also require that $\cancel{E}_T > 30$ GeV, and that $H_T > 250$ GeV when three or more jets are present. H_T is defined as the scalar sum of all of the transverse energy (lepton, jets, and \cancel{E}_T) in the event. Events are tagged using SLT_e tag level 1. The \cancel{E}_T and H_T requirements are optimized alongside the the SLT_e tag level. The optimization is described in more detail in Section 9.4. We designate the signal region for this analysis - where we expect an abundance of $t\bar{t}$ events - to be a events with three or more jets. The corresponding control region is those events with only one or two jets.

For convenience, the data is divided into data periods 0 through 11, corresponding to different time periods in data taking. Table 9.1 summarize the periods with the associated run numbers, dates, and integrated luminosity. CEM, CMUP, and CMX indicate leptons identified with the corresponding sub-detector. The entire CMX is unavailable before run 150145, and the keystone and miniskirts are only available beginning with Period 1. Since the silicon detector is an important component to conversion identification, we require that it is working properly. The total integrated luminosity up to March 2007 is approximately $1.7 \pm 0.1 \text{ fb}^{-1}$.

In total, through Period 11, we find 2196 events with 3 or more jets after the event selection described here. We then apply the Level 1 of the SLT_e tagger to this pretag sample. The selection criteria for SLT_e tags is summarized in Tables 4.6 and 4.7. We additionally require that the SLT_e tracks have a z_0 within 5 cm of the reconstructed vertex. We find 120 events in the signal region with one or more Level 1 SLT_e tags.

¹In general for this analysis, jet energies are corrected for detector effects, variations in η , multiple interactions, and the absolute energy scale. Jets are clustered using the JetClu0.4 algorithm. See Section 3.3.1 for more details.

Period	Run Range	Date	Integrated Luminosity (pb^{-1})	
			CEM/CMUP	CMX
0	138425 - 186598	04 Feb 02 - 22 Aug 04	332.87	319.47
1	190697 - 195408	07 Dec 04 - 18 Mar 05		
2	195409 - 198379	19 Mar 05 - 20 May 05		
3	198380 - 201349	21 May 05 - 19 Jul 05	363.01	363.01
4	201350 - 203799	20 Jul 05 - 04 Sep 05		
5	203819 - 206989	05 Sep 05 - 09 Nov 05		
6	206990 - 210011	10 Nov 05 - 14 Jan 06	258.37	258.37
7	210012 - 212133	14 Jan 06 - 22 Feb 06		
8	217990 - 222426	9 Jun 06 - 1 Sep 06	166.29	166.29
9	222529 - 228596	1 Sep 06 - 22 Nov 06	158.13	158.13
10	228664 - 233111	24 Nov 06 - 30 Jan 07	230.60	230.60
11	233133 - 237795	31 Jan 07 - 30 Mar 07	221.55	203.83

Table 9.1: Summary of the data run taking periods and their corresponding luminosity. The luminosity shown has a common 6% relative systematic uncertainty.

Pretag Event Yield by Data Period					
Period	1 Jet	2 Jets	3 Jets	4 Jets	≥ 5 Jets
0	22797	3657	250	125	39
1-4	25852	4317	297	127	37
5-7	18324	2962	178	99	32
8	11930	1937	145	62	22
9	11022	1748	139	55	17
10	15555	2606	159	98	32
11	15119	2468	190	79	14
Total	120599	19695	1358	645	193

Table 9.2: Pretag event yield collected from the high p_T lepton datasets. We require $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV.

Five of those events are tagged twice by the SLT $_e$. These results are summarized in Tables 9.2 and 9.3. A complete event list in the signal region is shown at the end of

Tag Event Yield by Run Range					
Period	1 Jet	2 Jets	3 Jets	4 Jets	≥ 5 Jets
0	229	81	9	8	4
1-4	278	92	14	8	2
5-7	212	68	6	11	3
8	139	39	3	3	3
9	134	38	5	3	3
10	167	58	11	8	4
11	153	51	8	4	0
Total	1312	427	56	45	19

Table 9.3: Tag event yield collected from the high p_T lepton datasets. We require $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV.

this chapter in Table 9.22.

9.2 Signal Expectation

We use Pythia MC with $M_{top} = 175$ GeV/ c^2 to simulate top pair production. We measure the signal acceptance by counting the number of events that pass the lepton+jets event selection described above divided by the total number of events generated. We do not restrict the decay channel at the generator level, so it is possible for some signal from other channels such as dilepton to be mis-categorized as lepton+jets. The measured acceptance is summarized in Table 9.4.

We then correct the acceptance with various scale factors to account for differences between MC modeling and data. These scale factors results from differences in the lepton identification and isolation as well as corrections for requirements imposed on data but not MC, such as the trigger efficiency, the muon track χ^2 requirement dis-

Uncorrected $t\bar{t}$ Acceptance (%)					
Lepton Type	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
CEM	0.176(2)	0.930(5)	1.515(6)	1.612(6)	0.560(4)
CMUP	0.109(2)	0.585(4)	0.966(5)	1.013(5)	0.348(3)
CMX Arches	0.040(1)	0.208(2)	0.331(3)	0.360(3)	0.122(2)
CMX Mini/Key	0.008(0)	0.044(1)	0.073(1)	0.077(1)	0.026(1)
Total	0.334(3)	1.767(6)	2.884(8)	3.062(8)	1.057(5)

Table 9.4: Uncorrected $t\bar{t}$ acceptance. We have required $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV. Statistical uncertainties are shown in parentheses.

cussed in Section 3.2.2, and the z vertex position. These corrections and scale factors are summarized in Tables 9.5 and 9.6. Table 9.7 shows the corrected acceptance.

We calculate the expected pretag yield from $t\bar{t}$ by multiplying the acceptance, \mathcal{A} , by the integrated luminosity and the expected cross section. The expected tag yield is calculated by weighting MC events by the probability to be tagged. This probability is calculated by applying the SLT_e tag matrix and fake matrix to all the taggable tracks in the event. If the track is flagged by the conversion filter, we rescale the probability by one minus the conversion (over-)efficiency SF for (non-)conversion electrons. The total probability to tag the event is then given by:

$$P_{event} = 1 - \prod_i (1 - p_i) \quad (9.1)$$

where p_i is the probability to tag a track. This is the same procedure used in Chapter 8. Naturally, we do not consider for tagging the track associated with the tight lepton.

Table 9.8 shows the pretag and tag expectations for $t\bar{t}$ assuming a 6.7 pb cross section, corresponding to $M_{top} = 175$ GeV/ c^2 . Note that this means that fake electron tags and conversion electron tags also contribute to the total signal expectation.

Correction	Period 0	Period 1-4	Period 5-7	Period 8
CEM				
Trigger Efficiency	0.962(7)	0.976(6)	0.979(4)	0.959(7)
Electron ID SF	0.991(5)	0.985(5)	0.974(4)	
Lepton Isolation		1.000(8)		
CMUP				
Trigger Efficiency	0.902(4)	0.919(4)	0.918(5)	0.913(6)
Muon ID SF	0.985(4)	0.989(4)	0.975(5)	0.975(6)
Muon Reconstruction	0.951(4)	0.939(4)	0.941(4)	0.955(5)
χ^2 Requirement		0.990(12)		
Lepton Isolation		1.000(10)		
CMX Arches				
Trigger Efficiency	0.967(4)	0.955(4)	0.954(5)	0.947(6)
Muon ID SF	1.014(4)	1.000(5)	1.004(6)	1.000(8)
Muon Reconstruction	0.996(2)	0.993(2)	0.989(3)	0.991(3)
χ^2 Requirement		0.989(17)		
Lepton Isolation		1.000(14)		
CMX Miniskirt/Keystone				
Trigger Efficiency	—	0.772(14)	0.744(19)	0.755(23)
Muon ID SF	—	0.979(11)	0.990(13)	1.001(15)
Muon Reconstruction	—	0.933(9)	0.939(11)	0.902(16)
χ^2 Requirement		0.989(17)		
Lepton Isolation		1.000(14)		
Common				
z Vertex < 60 cm		0.958(2)		
Isolated Track Efficiency		1.014(2)		

Table 9.5: A summary of scale factors and efficiencies used to correct the acceptance. Systematic uncertainties are shown in parentheses.

Approximately 25% of the tags in the $t\bar{t}$ sample are due to fake electrons and an additional 10% are due to conversion electrons.

Combined Lepton ID Scale Factor				
Data Period	CEM	CMUP	CMX	Miniskirt/Keystone
0	0.926(11)	0.813(14)	0.938(22)	—
1-4	0.934(11)	0.821(14)	0.911(21)	0.677(22)
5-7	0.926(10)	0.810(15)	0.910(22)	0.644(25)
8	0.907(11)	0.818(16)	0.902(22)	0.655(29)

Table 9.6: Combined lepton ID Scale Factor by lepton type and data period. These are the combined results from Table 9.5. Systematic uncertainties are shown in parentheses.

Corrected $t\bar{t}$ Acceptance (%)					
Lepton Type	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
CEM	0.163(3)	0.862(11)	1.403(17)	1.493(18)	0.519(7)
CMUP	0.089(2)	0.477(9)	0.788(15)	0.826(15)	0.284(6)
CMX (Arches)	0.037(1)	0.191(5)	0.304(8)	0.330(8)	0.112(3)
CMX (Mini/Key)	0.005(0)	0.029(1)	0.049(2)	0.051(2)	0.018(1)
Total	0.295(5)	1.559(24)	2.543(39)	2.700(41)	0.932(15)

Table 9.7: Corrected $t\bar{t}$ acceptance. We have required $H_T > 250$ GeV for events with ≥ 3 jets and $\cancel{E}_T > 30$ GeV. Combined statistical and systematic uncertainties are shown in parentheses.

9.3 Background Expectation

For this analysis, we use the ‘Method 2’ procedure of background estimation for the Lepton+Jets channel. The method divides the backgrounds into roughly three classes:

- MC backgrounds: These backgrounds have either a firm theoretical cross section or have a small enough contribution that the large uncertainty is unimportant. These backgrounds are treated in much the same way as we treat top, calculating

$t\bar{t}$ Expectation ($\sigma_{t\bar{t}} = 6.7 \text{ pb}, \int \mathcal{L} = 1.7 \text{ fb}^{-1}$)					
Expectation	1 Jet	2 Jets	3 Jets	4 Jets	≥ 5 Jets
Pretag	34.0 ± 2.1	180.5 ± 11.2	295.1 ± 18.3	313.3 ± 19.4	108.3 ± 6.7
Tag	1.2 ± 0.1	11.3 ± 1.0	22.4 ± 1.9	26.2 ± 2.2	10.6 ± 0.9

Table 9.8: Pretag and tag $t\bar{t}$ expectation assuming a top cross section of 6.7 pb . Uncertainties shown include the uncertainties associated with SLT_e tagging as well as lepton ID SFs, and luminosity.

an acceptance and efficiency and scaling by the cross section and luminosity.

For these backgrounds we consider diboson (WW, WZ, and ZZ), single top, Z+Jets, and Drell-Yan.

- QCD: This background results from multi-jet events where the primary lepton is faked or results from semi-leptonic decay of HF. The \cancel{E}_T signature results from the calorimeter resolution and/or soft neutrinos. We use the $\cancel{E}_T < 30$ region to calibrate the overall QCD contribution.
- W+Jets: This is the dominant background to this analysis. The Method 2 estimation of this background is calculated by assuming that the entire pretag sample – not already accounted for by $t\bar{t}$, MC backgrounds, and QCD – must be W+Jets. MC is then used to determine the total tag contribution from W+Jets given a pretag estimate.

9.3.1 MC Derived Backgrounds

For each of the MC derived backgrounds, we determine the pretag and tag contributions analogously to the $t\bar{t}$ contribution described in the previous section. For the

diboson background, we use Pythia MC. The single top MC is a Madevent/Pythia combination and is divided between s-channel and t-channel diagrams. We use the theoretical cross section scaled by 1/3 because the W boson has been forced to decay leptonically. Table 9.9 lists the samples and theoretical cross sections used.

Sample	Cross Section (pb)	Comments
single top (s)	0.88 ± 0.05	[26], MC requires leptonically decaying W
single top (t)	1.98 ± 0.08	"
WW	12.4 ± 0.25	[39]
WZ	3.96 ± 0.06	"
ZZ*	2.12 ± 0.15	"

Table 9.9: Cross section for various processes that contribute to the lepton+jets dataset.

Z+Jets MC has been generated with an Alpgen/Pythia combination. Z events can enter into the lepton+jets channel in a number of different ways. If the Z decays leptonically, only one lepton could be reconstructed properly, and one of the associated jets could be SLT_e tagged. Alternatively, both legs could be reconstructed, but the event still passes the Z veto and one leg is SLT_e tagged. Yet another possibility is that the Z decays to $\tau^+\tau^-$, and a τ decays leptonically yielding a tight lepton.

We use $Z(\rightarrow \ell^+\ell^-)+\text{Jets}$ MC including $Z(\rightarrow e/\mu/\tau)+b\bar{b}/c\bar{c}$. The total Z+jets $\cdot BF(Z \rightarrow \ell^+\ell^-)$ cross section calculated by Alpgen is approximately 184 pb, whereas the measured cross section is 264 ± 17 pb [40]. We scale the cross section by the ratio, 1.4, and then again by 1.2 ± 0.2 to account for the discrepancy in the prediction of the $N \geq 1$ jet distribution as shown in Figure 9.1. The Drell-Yan background is treated along the same lines, where we scale the theoretical cross section predicted by Alpgen by a factor of 1.68 ± 0.28 . We use the $DY(\rightarrow \ell^+\ell^-) + Np$ MC in three different off-shell Z

Sample	M_Z (GeV/ c^2)	Cross Section \times BF (pb)
$Z(\rightarrow \ell^+\ell^-) + 0p$	[75,105]	158
$Z(\rightarrow \ell^+\ell^-) + 1p$	[75,105]	21.6
$Z(\rightarrow \ell^+\ell^-) + 2p$	[75,105]	3.47
$Z(\rightarrow \ell^+\ell^-) + 3p$	[75,105]	0.550
$Z(\rightarrow \ell^+\ell^-) + \geq 4p$	[75,105]	0.0992
$Z(\rightarrow \ell^+\ell^-) + b\bar{b} + 0p$	[75,105]	0.511
$Z(\rightarrow \ell^+\ell^-) + b\bar{b} + 1p$	[75,105]	0.134
$Z(\rightarrow \ell^+\ell^-) + b\bar{b} + \geq 2p$	[75,105]	0.0385
$Z(\rightarrow \ell^+\ell^-) + c\bar{c} + 0p$	[75,105]	1.08
$Z(\rightarrow \ell^+\ell^-) + c\bar{c} + 1p$	[75,105]	0.332
$Z(\rightarrow \ell^+\ell^-) + c\bar{c} + \geq 2p$	[75,105]	0.107
<hr/>		
$DY(\rightarrow \ell^+\ell^-) + 0p$	[8,20]	1514
$DY(\rightarrow \ell^+\ell^-) + 1p$	[8,20]	19.7
$DY(\rightarrow \ell^+\ell^-) + \geq 2p$	[8,20]	6.98
$DY(\rightarrow \ell^+\ell^-) + 0p$	[20,75]	160
$DY(\rightarrow \ell^+\ell^-) + 1p$	[20,75]	8.39
$DY(\rightarrow \ell^+\ell^-) + 2p$	[20,75]	1.61
$DY(\rightarrow \ell^+\ell^-) + 3p$	[20,75]	0.233
$DY(\rightarrow \ell^+\ell^-) + \geq 4p$	[20,75]	0.0398
$DY(\rightarrow \ell^+\ell^-) + 0p$	[105,600]	4.07
$DY(\rightarrow \ell^+\ell^-) + 1p$	[105,600]	0.706
$DY(\rightarrow \ell^+\ell^-) + 2p$	[105,600]	0.117
$DY(\rightarrow \ell^+\ell^-) + 3p$	[105,600]	0.0185
$DY(\rightarrow \ell^+\ell^-) + \geq 4p$	[105,600]	0.00333

Table 9.10: Cross section times branching fraction ($\ell = e, \mu, \tau$) calculated by Alpgen for various Z or Drell-Yan processes. The branching fractions for each lepton are identical due to lepton universality. The cross sections shown are then scaled by 1.68 ± 0.28 to match the measured $Z + jets$ spectrum.

mass regions: $[8, 20]$ GeV/ c^2 , $[20, 75]$ GeV/ c^2 , and $[105, 600]$ GeV/ c^2 . The DY+HF contribution to this measurement is negligible. The theoretical cross sections, before scaling, are shown in Table 9.10.

We use a jet-based Alpgen overlap removal procedure which we will describe in Section 9.3.3 to fill the parton phase-space properly. This procedure removes a small fraction of events which are double counted in the various MC samples. Tables 9.11

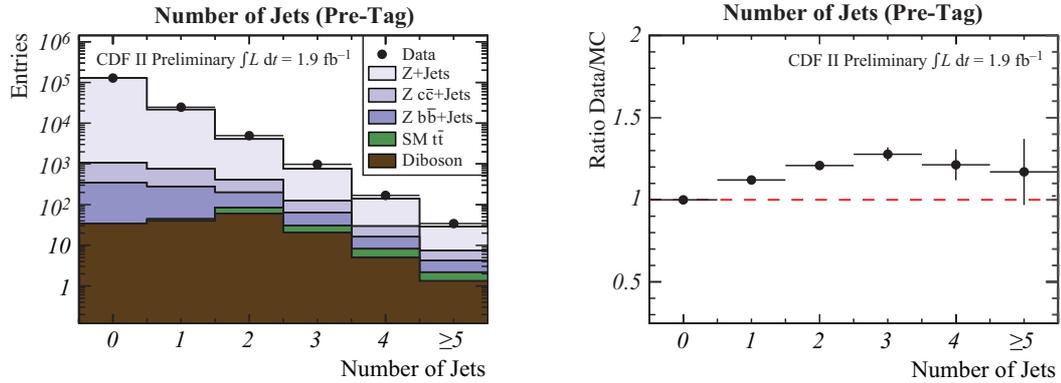


Figure 9.1: Jet multiplicity of measured and Alpgen predicted $Z \rightarrow \ell^+\ell^-$ events at CDF ($\ell = e, \mu$). The MC is normalized to the data, which is dominated by the 0 jet bin. On the left is a comparison of the distributions, and on the right is the ratio.

and 9.12 show the pretag and tag expectations for each of the MC backgrounds. The uncertainties shown include luminosity, lepton SF, cross sections, and SLT_e tagger uncertainties.

Pretag Composition for MC Backgrounds					
Process	1 jets	2 jets	3 jets	4 jets	≥ 5 jets
WW	524.92 ± 34.22	564.52 ± 36.79	44.89 ± 2.93	15.67 ± 1.02	3.91 ± 0.25
WZ	84.55 ± 5.40	102.62 ± 6.56	9.56 ± 0.61	3.23 ± 0.21	0.87 ± 0.06
ZZ	3.94 ± 0.37	3.67 ± 0.35	0.62 ± 0.06	0.30 ± 0.03	0.07 ± 0.01
single top (s)	17.18 ± 1.44	40.74 ± 3.42	7.10 ± 0.60	2.17 ± 0.18	0.53 ± 0.04
single top (t)	65.51 ± 4.85	72.30 ± 5.35	6.59 ± 0.49	1.37 ± 0.10	0.19 ± 0.01
Z+Jets	2660.69 ± 552.97	566.75 ± 107.32	46.99 ± 7.60	15.27 ± 2.50	3.51 ± 0.60
Drell-Yan	529.45 ± 110.95	127.93 ± 26.79	12.77 ± 2.67	3.22 ± 0.67	0.71 ± 0.15

Table 9.11: Lepton+jets pretag composition for MC derived backgrounds. Uncertainties shown include lepton ID systematics, luminosity, and cross section uncertainties.

Tag Composition for MC Backgrounds					
Process	1 jets	2 jets	3 jets	4 jets	≥ 5 jets
WW	12.87 ± 1.27	12.36 ± 1.14	1.53 ± 0.14	0.64 ± 0.06	0.25 ± 0.02
WZ	1.37 ± 0.13	3.04 ± 0.26	0.41 ± 0.04	0.21 ± 0.02	0.06 ± 0.01
ZZ	0.16 ± 0.02	0.17 ± 0.02	0.05 ± 0.01	0.02 ± 0.00	0.01 ± 0.00
Single Top (s)	0.55 ± 0.06	2.31 ± 0.23	0.46 ± 0.05	0.17 ± 0.02	0.05 ± 0.01
Single Top (t)	1.88 ± 0.17	2.67 ± 0.25	0.36 ± 0.03	0.09 ± 0.01	0.01 ± 0.00
Z +Jets	46.27 ± 10.14	19.52 ± 4.02	2.44 ± 0.44	1.09 ± 0.20	0.28 ± 0.05
Drell-Yan	10.01 ± 2.27	6.32 ± 1.42	1.11 ± 0.25	0.33 ± 0.07	0.09 ± 0.02

Table 9.12: Lepton+jets SLT_e Level 1 Tag composition for MC derived backgrounds. Uncertainties shown include lepton ID systematics, luminosity, and cross section uncertainties, as well as systematic uncertainties associated with the SLT_e tagger.

9.3.2 QCD

The QCD contribution to the pretag and tag samples is estimated by building \cancel{E}_T templates ranging from 0 to 120 GeV for each of the components that contribute to the lepton+jets channel and fitting them. Separate measurements are made for the 1, 2, and ≥ 3 jet bins, where we combine the full signal region due to limited statistics. The results of the fits are shown in Figures 9.2, 9.3, and 9.4.

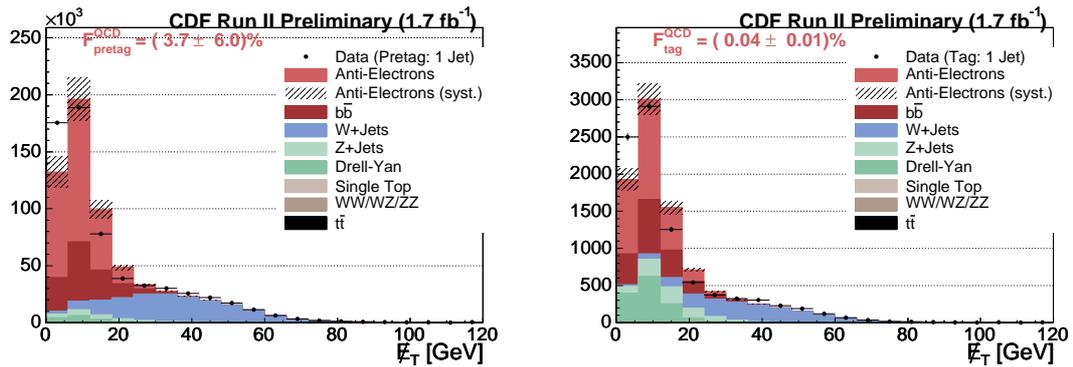


Figure 9.2: \cancel{E}_T fits showing the fraction of QCD in the pretag (left) and tag (right) regions of the 1 jet bin of the lepton+jets sample.

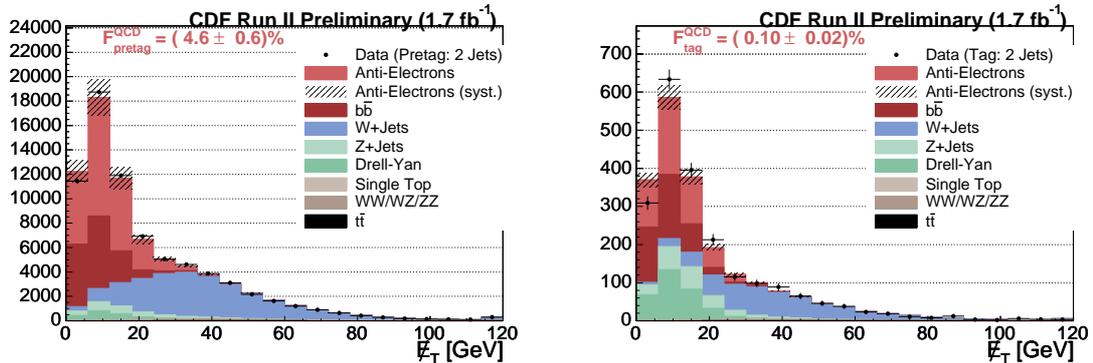


Figure 9.3: \cancel{E}_T fits showing the fraction of QCD in the pretag (left) and tag (right) regions of the 2 jet bin of the lepton+jets sample.

For most of the \cancel{E}_T templates, we use the MC predicted shape and normalize to the cross section. Separate shapes are calculated for the pretag and tag fits. This is done for each of the MC derived backgrounds (diboson, single top, Z +Jets, and Drell-Yan) as well as $t\bar{t}$ where we normalize to an assumed cross section of 6.7 pb. This assumption has a negligible effect on the cross section measurement. The W +Jets template is done in an identical way, except that we do not normalize to the cross section. Rather, we let the normalization float in the fit.

The QCD shape is estimated using two distinct components, a $b\bar{b}$ shape that we derive from MC, and an anti-electron shape that we derive from data. Anti-electrons are candidate central electrons with $E_T > 20$ GeV but which fail at least two standard electron requirements. Approximately 15% of these anti-electrons are real electrons. Because of limited statistics, the pretag anti-electron \cancel{E}_T template is also used to estimate the tag contribution. Both muons and electron QCD events are modeled in this way. Since we do not know the normalization for either of these backgrounds, we let the QCD normalizations float in the fit. In the ≥ 3 jet bin, the acceptance of

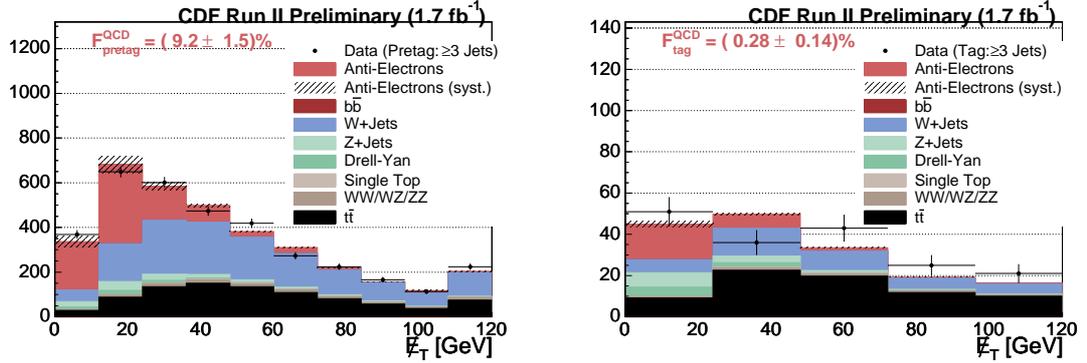


Figure 9.4: \cancel{E}_T fits showing the fraction of QCD in the pretag (left) and tag (right) regions of the 3 jet bin of the lepton+jets sample.

$b\bar{b}$ is so small that we set the total normalization of that template to 0.

The fit is performed by MINUIT which minimizes the χ^2 of the total background shape to the measured signal shape. If $\chi^2 > 1$, then we scale uncertainty on the fit by the χ^2 . Finally, we give the QCD shape an additional 15% systematic, based on the estimated purity of this sample.

To calculate the fraction of pretag events due to QCD, F_{pre}^{QCD} , we fit the pretag \cancel{E}_T distribution to the backgrounds letting the overall normalization of the anti-electrons, $b\bar{b}$ and W+Jets templates vary. F_{pre}^{QCD} then is the integral of the fitted anti-electron and $b\bar{b}$ shapes divided by the number of pretag events with $\cancel{E}_T > 30$ GeV. The tag fraction, F_{tag}^{QCD} , is similar, except the tag \cancel{E}_T templates are used, and we divide by the number of *pretag* events rather than the tag events. Therefore, the pretag estimation of QCD background is

$$N_{pre}^{QCD} = N_{pre}^{data} F_{pre}^{QCD} \quad (9.2)$$

	1 Jet	2 Jets	≥ 3 Jets
F_{pre}^{QCD} (%)	3.7 ± 6.0	4.6 ± 0.6	9.2 ± 1.5
F_{tag}^{QCD} (%)	0.045 ± 0.011	0.10 ± 0.02	0.28 ± 0.14

Table 9.13: Summary of the fraction of tag and pretag events in the lepton+jets channel due to QCD. These values have been derived from fits of the \cancel{E}_T in each of the jet bins.

and the tag estimation of the QCD background is

$$N_{tag}^{QCD} = N_{pre}^{data} F_{tag}^{QCD}. \quad (9.3)$$

Table 9.13 lists the results in each jet bin and for tag and pretag fractions. The uncertainties shown include not only the fit uncertainties but additionally the uncertainties associated with the lepton ID SF, cross sections, luminosity, and SLT_e tagging systematics. The additional uncertainties are calculated by varying the overall normalization of each fixed background up and down according to the relative uncertainty of the given systematic. For example, we scale each of the fixed backgrounds +6% and -6% relative and refit to account for the uncertainty on F_{QCD} due to the luminosity.

9.3.3 W+Jets

Once we have accounted for the MC and QCD backgrounds, the Method 2 assumption is that what remains is only W +Jets and $t\bar{t}$. Although we do not know the precise amount of $t\bar{t}$ in the sample before we make the measurement, we can use the formula

$$N_{pre}^W = N_{pre}^{data} (1 - F_{pre}^{QCD}) - \mathcal{A}_{t\bar{t}} \sigma_{t\bar{t}} \int \mathcal{L} dt - N_{pre}^{diboson} - N_{pre}^{singletop} - N_{pre}^{Z/DY} \quad (9.4)$$

to determine the pretag contribution of the W +Jets sample. We will address how we calculate $\sigma_{t\bar{t}}$ in Section 9.5. For now, we note that we have calculated every other variable in this equation.

To solve for the tag contribution from W +jets, we use the formula

$$N_{tag}^W = N_{pre}^W \epsilon_W \quad (9.5)$$

$$N_{tag}^W = N_{pre}^W \left(\sum_i \epsilon_i^{HF} F_i^{HF} + \epsilon^{LF} (1 - \sum_i F_i^{HF}) \right) \quad (9.6)$$

where F_i^{HF} are the Heavy Flavor fractions, and ϵ_i^{HF} and ϵ^{LF} are the efficiencies to tag HF and LF, respectively. i varies over the number of observed HF jets in the event, forced into the categories $1b$, $2b$, $1c$, or $2c$. The calculation of each of these quantities follows.

This approach is motivated by the fact that there are large uncertainties on the total W +jets cross section as well as the fraction of W +jets which contains HF. This method allows us to take the overall normalization of the W +jets sample from data as well as calibrate the HF fraction of W +jets in data.

HF Fraction Estimation and Calibration

The HF fractions are the fractions of W +jet events that contain b and c jets in the final state. These fractions are determined from MC, but calibrated with generic

jet data. The fractions are calculated according to the formulas

$$w^i = \frac{\sigma^i}{N_{gen}^i} \quad (9.7)$$

$$F_{2b} = \frac{\sum w^i N_{\geq 2b}^i}{\sum w^i N^i} \quad (9.8)$$

$$F_{1b} = \frac{\sum w^i N_{1b}^i}{\sum w^i N^i} \quad (9.9)$$

$$F_{2c} = \frac{\sum w^i N_{\geq 2c,0b}^i}{\sum w^i N^i} \quad (9.10)$$

$$F_{1c} = \frac{\sum w^i N_{1c,0b}^i}{\sum w^i N^i} \quad (9.11)$$

where σ is the cross section N is the number of events generated, and i runs over the various W +jets MC datasets. The subscripts of N indicate events with a specific number of jets matched ($\Delta R \leq 0.4$) to a bottom or charm hadron in the MC. A summary of the cross section times branch fraction of these processes is given in Table 9.14. Note that the cross sections are used only as weights and that the overall normalization comes directly from data. W +jet MC events have overlap events removed, a procedure described Section 9.3.3.

The HF fractions that are calculated are then scaled by a correction - K factor - which accounts for the mis-measurement of the HF fractions in MC. The K factor is estimated to be 1.0 ± 0.4 by comparing the vertex mass and pseudo- $c\tau$ shapes of SECVTX tagged jets in dijet data and MC. These shapes determine relative fractions of bottom, charm, and light jets in comparable dijet samples. Uncertainties on the K factor include variations of the value in jet multiplicity as well as the Q^2 scale of the generated MC. The K factor is not applied to W +charm events, but a 30% systematic dominated by the Q^2 scale is still applied. This procedure is described more fully in [41]. After the K factor correction, we measure HF fractions shown in Table 9.15.

Sample	Cross Section×BF (pb)
$W(\rightarrow \ell\nu) + 0p$	181×10^3
$W(\rightarrow \ell\nu) + 1p$	225
$W(\rightarrow \ell\nu) + 2p$	35.3
$W(\rightarrow \ell\nu) + 3p$	5.59
$W(\rightarrow \ell\nu) + \geq 4p$	1.03
$W(\rightarrow \ell\nu) + b\bar{b} + 0p$	2.98
$W(\rightarrow \ell\nu) + b\bar{b} + 1p$	0.888
$W(\rightarrow \ell\nu) + b\bar{b} + \geq 2p$	0.287
$W(\rightarrow \ell\nu) + c\bar{c} + 0p$	5.00
$W(\rightarrow \ell\nu) + c\bar{c} + 1p$	1.79
$W(\rightarrow \ell\nu) + c\bar{c} + \geq 2p$	0.628
$W(\rightarrow \ell\nu) + c + 0p$	17.1
$W(\rightarrow \ell\nu) + c + 1p$	3.39
$W(\rightarrow \ell\nu) + c + 2p$	0.507
$W(\rightarrow \ell\nu) + c + \geq 3p$	0.083

Table 9.14: Cross section times branching fraction ($\ell = e, \mu, \tau$) calculated by Alpgen for various W processes. The branching fractions for each lepton are identical due to lepton universality.

Like the HF fractions, the efficiencies - defined as the average SLT_e event tagging probability, P_{event} (see Equation 9.1) - are calculated in events with a specific number of jets matched to bottom and charm hadrons. The light flavor efficiency is measured in those events with no HF hadrons matched to jets. The values for the measured efficiencies are shown in Table 9.16.

Overlap Removal Scheme

We use a combination of Alpgen and Pythia MC to simulate W/Z /Drell-Yan events, where Alpgen is responsible for calculating the matrix elements, and Pythia for showering and fragmenting the partons. However, this distinction can be ambiguous. For instance, a $W + b\bar{b}$ event in the final state could be produced two different ways:

Fraction	1 jets	2 jets	3 jets	≥ 4 jets
F_{1b}	0.8 ± 0.3	1.6 ± 0.6	3.0 ± 1.1	3.7 ± 1.4
F_{2b}	—	1.0 ± 0.4	2.2 ± 0.8	3.5 ± 1.3
F_{1c}	5.8 ± 1.6	9.1 ± 2.6	10.2 ± 3.3	12.1 ± 3.9
F_{2c}	—	1.5 ± 0.6	3.4 ± 1.3	6.3 ± 2.3

Table 9.15: Heavy Flavor Fractions for W+Jet events. Uncertainties are systematic only, related to the K factor and the Q^2 scale of the generated MC. All numbers are shown in units of %.

HF Jets	1 jets	2 jets	3 jets	≥ 4 jets
0 b, 0 c	0.92 ± 0.06	1.89 ± 0.11	3.01 ± 0.17	4.24 ± 0.24
2 b	—	6.72 ± 0.33	7.26 ± 0.37	9.55 ± 0.45
1 b	3.33 ± 0.16	4.39 ± 0.22	5.43 ± 0.29	6.80 ± 0.36
2 c	—	3.11 ± 0.17	4.17 ± 0.23	5.58 ± 0.30
1 c	1.61 ± 0.09	2.50 ± 0.14	3.46 ± 0.20	4.78 ± 0.28

Table 9.16: SLT_e Level 1 tagging efficiency for different classes of HF in W+Jet events. Uncertainties shown include all SLT_e tagging systematics. All numbers are shown in units of %.

- Alpgen generates a $W + b\bar{b}$ event
- Alpgen generates a $W + 1p$ event, where the p is a gluon and Pythia splits the gluon into a $b\bar{b}$ pair

Sherman [41] discusses various approaches to this problem. The solution we use is to select HF production from the matrix element or shower through a jet-based matching procedure. We allow Alpgen to predict the phase-space characterized by large opening angles between heavy quarks, but let Pythia account for smaller opening angles where the showering algorithm has been tuned. We reject events in our MC samples when the matrix element quarks are matched ($\Delta R \leq 0.4$) to the same reconstructed jet. We also reject events when the HF quarks from the shower do not match to the same reconstructed jet.

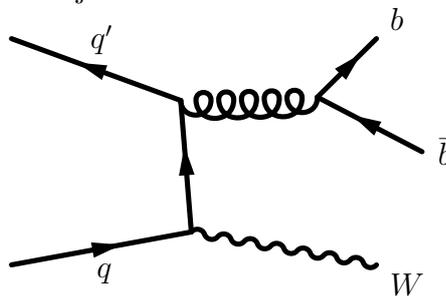


Figure 9.5: Feynman diagram of $W + b\bar{b}$ event. This can be generated as a $W + b\bar{b}$ event by Alpgen or a $W + 1p$ event by Alpgen with a subsequent gluon split by Pythia.

9.4 Optimization

We chose to optimize the H_T and \cancel{E}_T requirements and SLT_e tag level choice simultaneously on the expected uncertainty. This was done assuming a $t\bar{t}$ cross section of 6.7 pb. However, since the fitted value of F_{tag}^{QCD} depends strongly on the tagged

data, we cannot use the fit in the calculation of the expected uncertainty. Instead we estimate the value as $F_{tag}^{QCD} = 0.03F_{pre}^{QCD}$ and neglect the fit uncertainty. We scanned H_T in increments of 10 GeV and \cancel{E}_T in increments of 5 GeV. We considered all three tag levels. With this method we found that the expected uncertainty was minimized with requirements on $H_T > 250$ GeV for ≥ 3 jets, $\cancel{E}_T > 30$ GeV, and tag level 1.

The expected cross section with $H_T > 250$ GeV and $\cancel{E}_T > 30$ GeV, with SLT_e tag level 1 is

$$\sigma_{t\bar{t}}(\text{pb}) = 6.7 \pm 2.3(\text{stat}) \pm 1.5(\text{syst}) \pm 0.5(\text{lumi}). \quad (9.12)$$

For tag level 2, we expect

$$\sigma_{t\bar{t}}(\text{pb}) = 6.7 \pm 2.5(\text{stat}) \pm 1.5(\text{syst}) \pm 0.5(\text{lumi}). \quad (9.13)$$

For tag level 3, we expect

$$\sigma_{t\bar{t}}(\text{pb}) = 6.7 \pm 2.8(\text{stat}) \pm 1.6(\text{syst}) \pm 0.5(\text{lumi}). \quad (9.14)$$

9.5 Measurement

In the preceding sections we have calculated the various elements of the $t\bar{t}$ cross section calculation. Although the background is a function of the cross section, we can remove that dependence with a little algebraic manipulation. We find that

$$\sigma_{t\bar{t}} = \frac{(N_{pre}^{data}(F_{pre}^{QCD} - 1) + N_{pre}^{MC})\epsilon_W + N_{tag}^{data} - N_{pre}^{data}F_{tag}^{QCD} - N_{tag}^{MC}}{(\epsilon_{t\bar{t}} - \epsilon_W)\mathcal{A}_{t\bar{t}} \int \mathcal{L} dt} \quad (9.15)$$

where ϵ_W is the effective W tagging efficiency (see Equation 9.6) and $N_{pre/tag}^{MC}$ is the total number of MC background events either pretag or tag. This assumes that the QCD fraction has little to no dependence on the top cross section, which is true if

QCD contribution to the signal region is small. By using this formula, we can then calculate the statistical and systematic uncertainties analytically.

Solving for the cross section, we find that $\sigma_{t\bar{t}} = 8.0 \pm 2.4(\text{stat})$ pb. The statistical uncertainty has been verified by running pseudo-experiments. In table 9.17, we show the sample composition with the measured 8.0 pb $t\bar{t}$ cross section. This table is represented graphically in Figure 9.6.

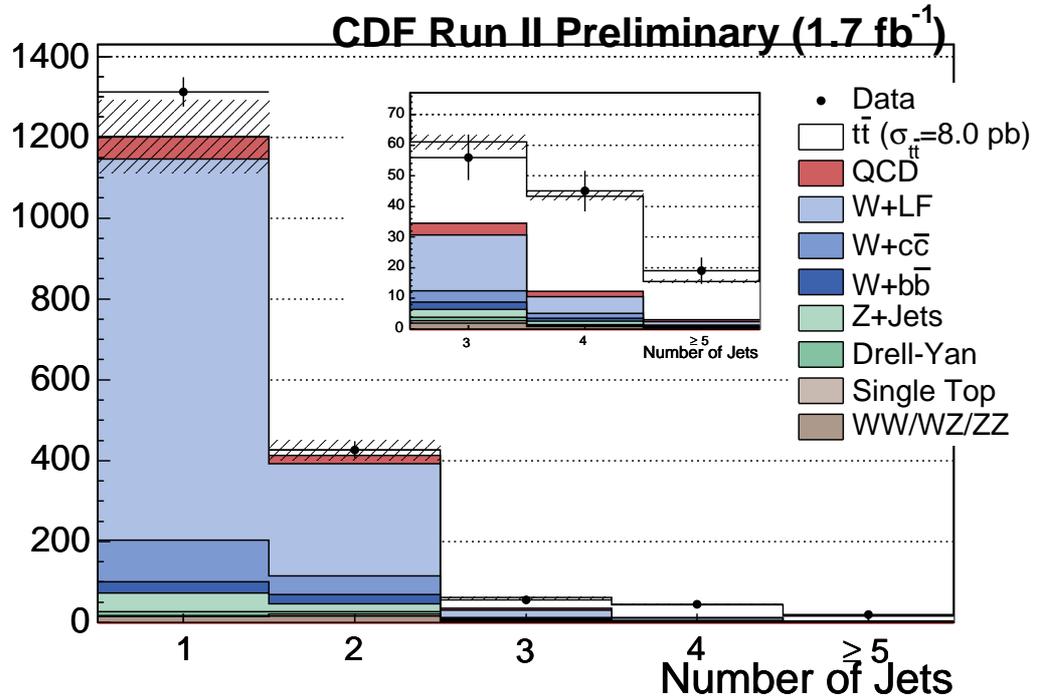


Figure 9.6: Jet multiplicity of SLT_e tagged events in the lepton+jets dataset. The embedded plot is the ≥ 3 jet subsample.

Although we have explicitly accounted for most of the systematics in this analysis already, a few still remain. We calculate the uncertainty of the cross section on the jet energy scale (JES) by adjusting the jet energy correction $\pm 1\sigma$ and measuring the

CDF Run II Preliminary (1.7 fb^{-1})					
Process	1 jets	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	120599	19695	1358	645	193
Pretag $t\bar{t}$ ($\sigma=7.84 \text{ pb}$)	39.82 ± 2.11	211.20 ± 11.18	345.35 ± 18.29	366.64 ± 19.42	126.67 ± 6.71
WW	12.87 ± 1.27	12.36 ± 1.14	1.53 ± 0.14	0.64 ± 0.06	0.25 ± 0.02
WZ	1.37 ± 0.13	3.04 ± 0.26	0.41 ± 0.04	0.21 ± 0.02	0.06 ± 0.01
ZZ	0.16 ± 0.02	0.17 ± 0.02	0.05 ± 0.01	0.02 ± 0.00	0.01 ± 0.00
Single Top (s)	0.55 ± 0.06	2.31 ± 0.23	0.46 ± 0.05	0.17 ± 0.02	0.05 ± 0.01
Single Top (t)	1.88 ± 0.17	2.67 ± 0.25	0.36 ± 0.03	0.09 ± 0.01	0.01 ± 0.00
Z +Jets	46.27 ± 10.14	19.52 ± 4.02	2.44 ± 0.44	1.09 ± 0.20	0.28 ± 0.05
Drell-Yan	10.01 ± 2.27	6.32 ± 1.42	1.11 ± 0.25	0.33 ± 0.07	0.09 ± 0.02
QCD	53.87 ± 14.14	20.20 ± 4.65	3.75 ± 1.92	1.78 ± 0.91	0.53 ± 0.27
$W + b\bar{b}$	28.18 ± 10.94	22.74 ± 8.70	2.43 ± 0.94	1.04 ± 0.43	0.23 ± 0.10
$W + c\bar{c}, W + c$	104.21 ± 30.19	47.09 ± 14.61	3.80 ± 1.31	1.66 ± 0.62	0.36 ± 0.15
W +Light Flavor	960.67 ± 90.79	280.97 ± 22.86	18.56 ± 2.10	5.60 ± 1.14	1.22 ± 0.32
Total W +Jets	1093.06 ± 101.27	350.80 ± 24.04	24.78 ± 2.05	8.30 ± 1.38	1.81 ± 0.43
Background	1220.04 ± 94.80	417.39 ± 25.50	34.89 ± 2.36	12.64 ± 1.32	3.09 ± 0.41
$t\bar{t}$ ($\sigma=7.84 \text{ pb}$)	1.41 ± 0.10	13.25 ± 0.96	26.27 ± 1.94	30.70 ± 2.16	12.41 ± 0.86
Tags	1314	432	56	45	20

Table 9.17: Sample Composition with SLT_e tag level 1. Uncertainties include luminosity, lepton SFs, cross section uncertainties, tagger systematics, K-factor, and the QCD fit.

CDF Run II Preliminary (1.7 fb^{-1})					
Process	1 jets	2 jets	3 jets	4 jets	≥ 5 jets
	120599	19695	1358	645	193
Pretag $t\bar{t}$ ($\sigma=7.82 \text{ pb}$)	39.71 ± 2.11	210.60 ± 11.18	344.38 ± 18.29	365.60 ± 19.42	126.31 ± 6.71
WW	10.57 ± 1.07	9.63 ± 0.89	1.18 ± 0.11	0.49 ± 0.04	0.19 ± 0.02
WZ	1.11 ± 0.10	2.43 ± 0.21	0.33 ± 0.03	0.17 ± 0.01	0.05 ± 0.00
ZZ	0.13 ± 0.02	0.14 ± 0.02	0.04 ± 0.00	0.02 ± 0.00	0.01 ± 0.00
Single Top (s)	0.45 ± 0.05	1.89 ± 0.19	0.37 ± 0.04	0.14 ± 0.01	0.04 ± 0.00
Single Top (t)	1.54 ± 0.14	2.16 ± 0.20	0.29 ± 0.03	0.07 ± 0.01	0.01 ± 0.00
Z +Jets	37.37 ± 8.22	15.73 ± 3.26	1.96 ± 0.36	0.88 ± 0.16	0.22 ± 0.04
Drell-Yan	8.16 ± 1.86	5.15 ± 1.17	0.91 ± 0.20	0.27 ± 0.06	0.07 ± 0.02
QCD	40.18 ± 10.30	14.52 ± 3.48	2.06 ± 1.58	0.98 ± 0.75	0.29 ± 0.22
$W + b\bar{b}$	23.19 ± 9.00	18.56 ± 7.11	1.96 ± 0.76	0.84 ± 0.35	0.18 ± 0.08
$W + c\bar{c}, W + c$	83.08 ± 24.08	36.91 ± 11.47	2.95 ± 1.02	1.29 ± 0.49	0.28 ± 0.12
W +Light Flavor	725.28 ± 68.65	212.37 ± 17.32	14.08 ± 1.60	4.27 ± 0.87	0.93 ± 0.25
Total W +Jets	831.54 ± 77.50	267.85 ± 18.75	19.00 ± 1.60	6.41 ± 1.07	1.40 ± 0.33
Background	931.05 ± 73.15	319.50 ± 19.90	26.14 ± 1.90	9.43 ± 1.04	2.29 ± 0.32
$t\bar{t}$ ($\sigma=7.82 \text{ pb}$)	1.16 ± 0.08	10.84 ± 0.80	21.39 ± 1.62	24.78 ± 1.78	9.97 ± 0.71
Tags	995	337	42	36	16

Table 9.18: Sample Composition with SLT_e tag level 2. Uncertainties include luminosity, lepton SFs, cross section uncertainties, tagger systematics, K-factor, and the QCD fit.

CDF Run II Preliminary (1.7 fb^{-1})					
Process	1 jets	2 jets	3 jets	4 jets	≥ 5 jets
Pretag	120599	19695	1358	645	193
Pretag $t\bar{t}$ ($\sigma=5.88 \text{ pb}$)	29.86 ± 2.11	158.37 ± 11.18	258.97 ± 18.29	274.93 ± 19.42	94.99 ± 6.71
WW	7.32 ± 0.82	6.22 ± 0.62	0.76 ± 0.08	0.30 ± 0.03	0.12 ± 0.01
WZ	0.75 ± 0.08	1.63 ± 0.16	0.22 ± 0.02	0.11 ± 0.01	0.03 ± 0.00
ZZ	0.09 ± 0.01	0.10 ± 0.01	0.03 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Single Top (s)	0.31 ± 0.03	1.31 ± 0.15	0.25 ± 0.03	0.09 ± 0.01	0.03 ± 0.00
Single Top (t)	1.06 ± 0.11	1.47 ± 0.15	0.20 ± 0.02	0.05 ± 0.00	0.01 ± 0.00
Z +Jets	25.30 ± 5.68	10.62 ± 2.26	1.32 ± 0.25	0.60 ± 0.11	0.15 ± 0.03
Drell-Yan	5.59 ± 1.31	3.53 ± 0.82	0.63 ± 0.14	0.19 ± 0.04	0.05 ± 0.01
QCD	24.62 ± 5.71	7.88 ± 2.38	2.54 ± 1.35	1.21 ± 0.64	0.36 ± 0.19
$W + b\bar{b}$	16.10 ± 6.29	12.80 ± 4.93	1.47 ± 0.57	0.86 ± 0.34	0.22 ± 0.09
$W + c\bar{c}, W + c$	55.36 ± 16.19	24.09 ± 7.54	2.11 ± 0.73	1.25 ± 0.44	0.33 ± 0.12
W +Light Flavor	446.19 ± 44.89	131.12 ± 11.56	9.67 ± 1.11	3.98 ± 0.64	1.04 ± 0.18
Total W +Jets	517.64 ± 51.73	168.01 ± 13.40	13.25 ± 1.17	6.08 ± 0.70	1.59 ± 0.21
Background	582.68 ± 50.29	200.77 ± 14.34	19.20 ± 1.59	8.64 ± 0.79	2.35 ± 0.23
$t\bar{t}$ ($\sigma=5.88 \text{ pb}$)	0.60 ± 0.07	5.64 ± 0.63	11.06 ± 1.25	12.68 ± 1.36	5.06 ± 0.53
Tags	597	226	25	23	11

Table 9.19: Sample Composition with SLT_e tag level 3. Uncertainties include luminosity, lepton SFs, cross section uncertainties, tagger systematics, K-factor, and the QCD fit.

effect on the pretag and tag content as well as \cancel{E}_T shapes of all the MC samples. We measure a cross section of 7.2 with $+1\sigma$ JES and 8.5 with -1σ JES. Therefore we assign an 8.6% relative systematic on the JES.

We also calculate the effect of initial state radiation (ISR), final state radiation (FSR), different PDFs, as well as MC modeling on the top acceptance. To calculate the effect of ISR and FSR, we remeasure the acceptance using MC with the ISR and FSR either doubled or halved. We take the mean deviation as a systematic uncertainty. We also take the difference in the $t\bar{t}$ acceptance between Herwig and Pythia MC generators as a systematic, making sure to adjust for the different W branching fraction. This systematic is intended to capture differences in jet fragmentation models. The uncertainty due to PDFs are calculated by rerunning over a $t\bar{t}$ MC dataset generated with a different set of PDFs. The total systematics are shown in Table 9.21.

Therefore, we calculate a $t\bar{t}$ cross section with tag level 1 of:

$$\sigma_{t\bar{t}}(\text{pb}) = 7.8 \pm 2.4(\text{stat}) \pm 1.6(\text{syst}) \pm 0.5(\text{lumi}). \quad (9.16)$$

The cross section measured with tag level 2 is

$$\sigma_{t\bar{t}}(\text{pb}) = 7.8 \pm 2.5(\text{stat}) \pm 1.6(\text{syst}) \pm 0.5(\text{lumi}). \quad (9.17)$$

Tag level 3 results in

$$\sigma_{t\bar{t}}(\text{pb}) = 5.9 \pm 2.7(\text{stat}) \pm 1.6(\text{syst}) \pm 0.5(\text{lumi}). \quad (9.18)$$

If we divide the events into exclusive channels (*i.e.* tag level 1, but not 2 or 3), we can pinpoint disagreement in tag level 3. Table 9.20 shows the measured and expected

tags by exclusive tag level. Assuming a cross section of 7.84 pb, we see that whereas the expected and measured tags agree for exclusive tag level 1, exclusive tag level 2 has five too many tags, and exclusive tag level 3 has 5 too few. Since $\sqrt{35} \approx 5.9$, we consider this within $\pm 1\sigma$ of the statistical uncertainties.

Tag Level	Exp. Background	Exp. $t\bar{t}$	Total Expected	Measured
1 ($\sigma_{t\bar{t}} = 7.84$ pb)	12.8	13.0	25.8	26
2 ($\sigma_{t\bar{t}} = 7.84$ pb)	11.7	18.0	29.7	35
3 ($\sigma_{t\bar{t}} = 7.84$ pb)	26.1	38.4	64.5	59
3 ($\sigma_{t\bar{t}} = 5.88$ pb)	30.2	28.8	59.0	59

Table 9.20: Comparison of measured and expected tags in the signal region (≥ 3 jets) of the lepton+jets channel. SLT_e Tag levels shown are exclusive and assume $\sigma_{t\bar{t}} = 8.0$ pb unless otherwise indicated.

Another point worth noting here is that the dominant uncertainties are associated with the tagger despite that the typical uncertainty on the event tagging probability is around 5% relative. The reason for this discrepancy can be readily seen from Equation 9.15. The denominator is proportional to $1/(\epsilon_{t\bar{t}} - \epsilon_W)$. When this portion of the denominator is small, the uncertainties on the efficiency (dominated by the tagger systematics) become magnified.

The remaining figures in this chapter show kinematic distributions from the signal region.

Systematic	Relative Uncertainty on $\sigma_{t\bar{t}}$ (%)
Jet Energy Scale	8.4
Non-W Fit	5.0
K-Factor	3.0
Herwig	2.2
Lepton ID	1.6
Background Cross Section	0.6
PDF	0.9
FSR	0.6
ISR	0.5
Conversion (over-)efficiency	10.7
Fake Matrix	7.8
Calorimeter Modeling	7.7
Tag Matrix	6.8
Jet Environment correction	5.4
Total tagger syst.	17.6
Total	20.6

Table 9.21: Summary of systematic uncertainties on the $t\bar{t}$ cross section measurement. Uncertainties shown are relative.

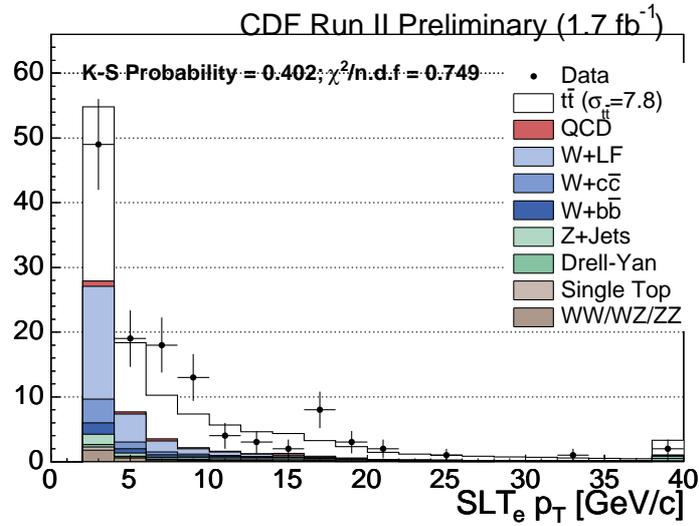


Figure 9.7: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track p_T . tag track.

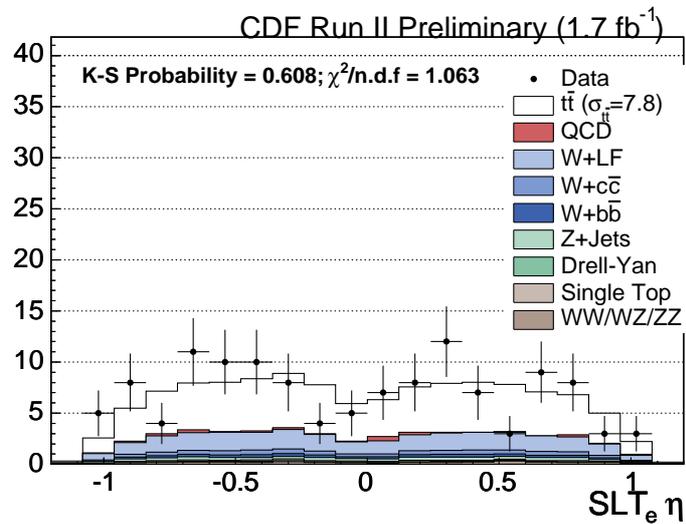


Figure 9.8: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track η . tag track.

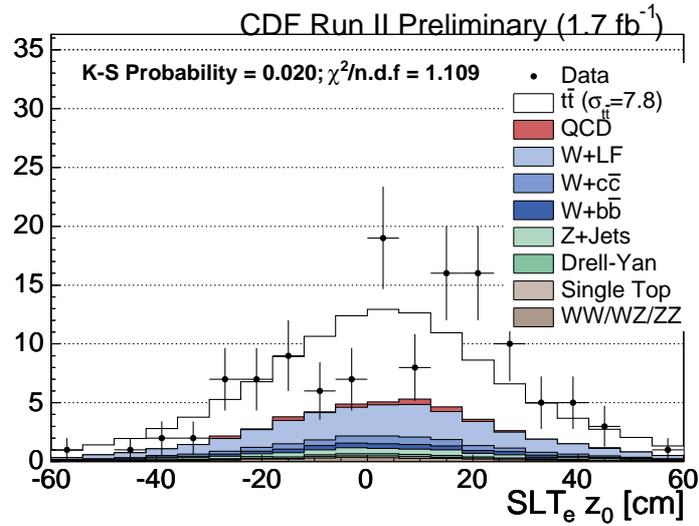


Figure 9.9: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track z_0 . tag track.

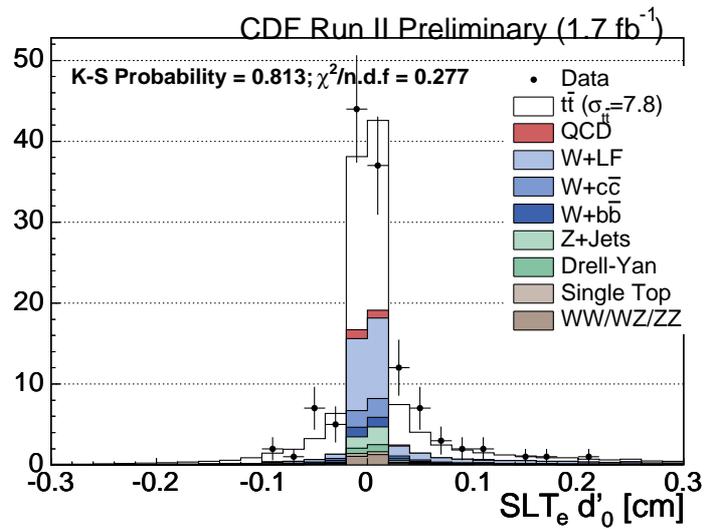


Figure 9.10: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track d'_0 . tag track.

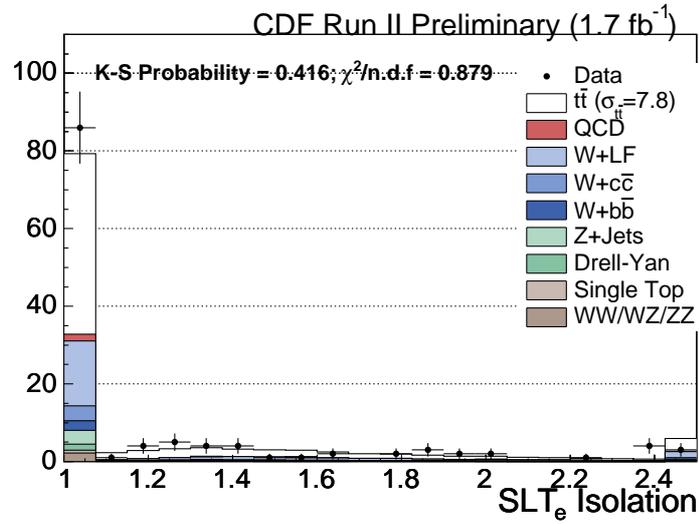


Figure 9.11: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the track isolation. tag track.

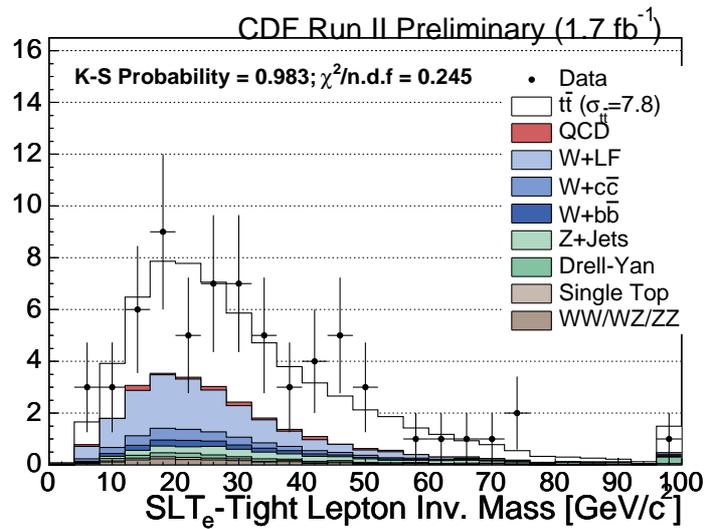


Figure 9.12: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the invariant mass between the SLT_e tag and the tight W lepton. tag track.

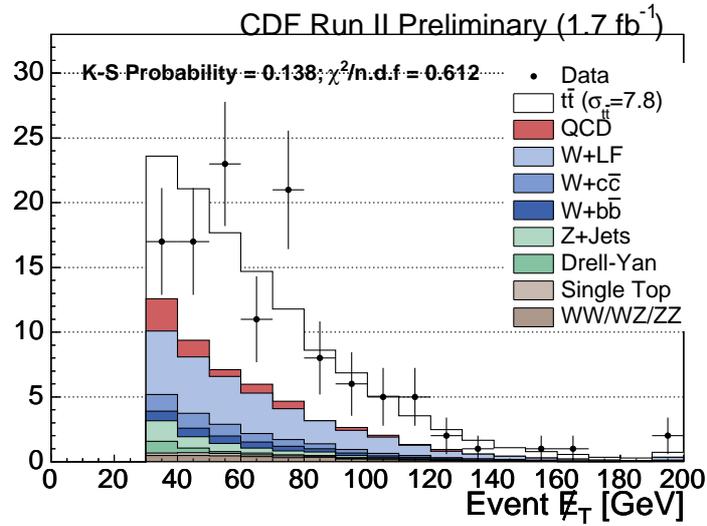


Figure 9.13: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the event E_T . tag track.

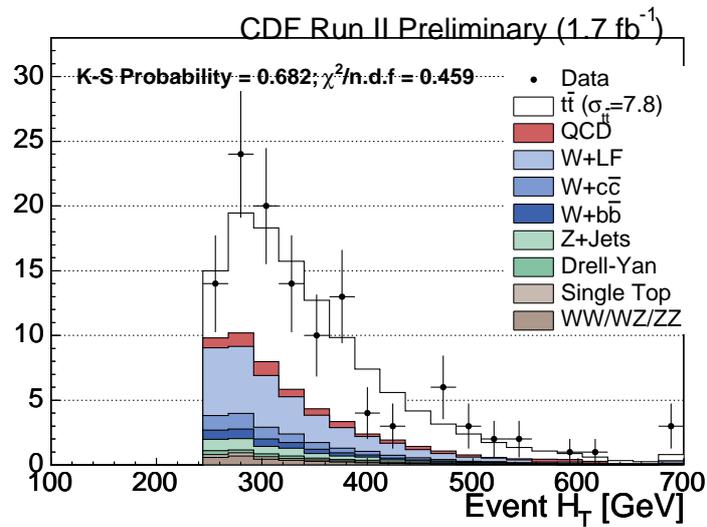


Figure 9.14: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Show is the event H_T . tag track.

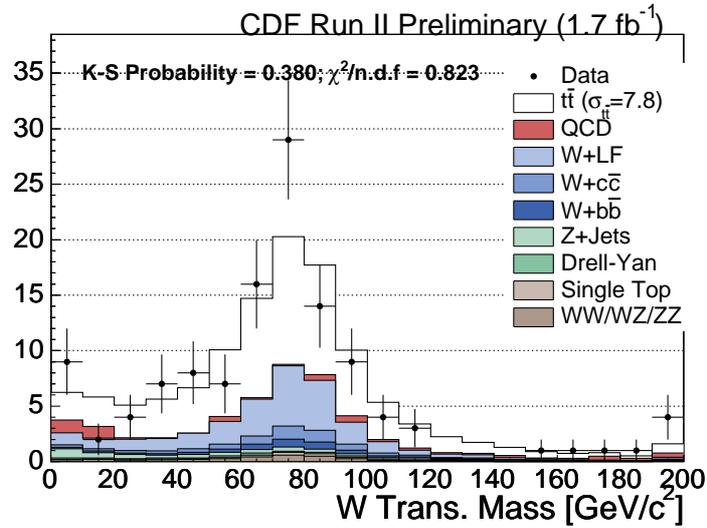


Figure 9.15: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Shows the reconstructed transverse mass of the W tag track.

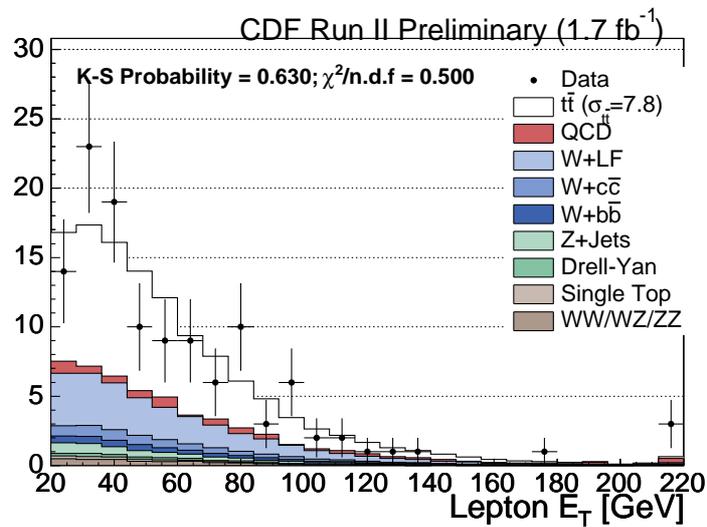


Figure 9.16: Kinematic distributions of the Level 1 SLT_e tags in the signal region. The top signal is scaled to the measured value. Shows the lepton p_T/E_T tag track.

Table 9.22: List of events in the lepton+jets signal region with ≥ 1 SLT_e Tag. Shown is the run number, event number, the number of jets in the event, the lepton type and charge, the tag level and charge, and the corresponding p_T of the SLT_e tagged track. An asterisk indicates the track was also in a loose SECVTX tagged jet. Multiple entries for the tag information indicate a double tag. Five events have two SLT_e tags.

Run	Event	Jets	Lepton (Charge)	Tag Level (Charge)	P_T (GeV/c)
148153	6088	3	CEM(-)	2(+)	17.3675
153374	730854	3	CEM(+)	3(+)	8.9373
156006	681024	3	CEM(-)	1(-)	2.39108
160230	3976545	3	CMX(-)	3(+)	6.62551
161414	68227	5	CEM(-)	2(-)	5.17401
161678	1080224	4	CEM(-)	3(-), 3(+)	2.72737, 2.40812
162631	163651	5	CEM(-)	3(+)	4.64119
166529	4938	4	CMUP(-)	*1(-)	2.95741
167551	3626393	5	CMUP(+)	*2(-)	4.33914
167955	2856469	4	CMUP(-)	1(+)	18.0613
168001	421782	3	CMUP(-)	3(+)	2.00958
168563	2395692	4	CEM(-)	2(+)	21.622
177927	3216152	3	CEM(+)	*3(-)	3.69241
178390	3929243	3	CMUP(+)	1(+)	3.45154
178761	1716435	5	CEM(-)	3(-)	9.78115
182874	765090	4	CEM(+)	3(+)	2.02286
183125	1590533	3	CEM(-)	*2(+)	2.75218
183557	1288983	3	CMUP(-)	*3(+)	5.41405
184453	19917	4	CEM(+)	*2(+)	7.53854
185332	4430084	4	CMUP(+)	*3(-), *3(+)	16.9809, 3.08167
186041	498698	4	CMX(-)	2(+)	4.05235
191766	1166089	3	CEM(+)	*1(+)	5.40545
193892	367844	4	CMX(-)	2(-)	2.1932
194147	5524014	3	CMUP(+)	2(-)	2.12592
196085	1648156	4	CMUP(-)	*3(+)	16.264
196099	10495653	3	CMUP(-)	3(-)	9.76928
196441	5763852	4	CEM(-)	1(+)	4.2793
196489	166456	5	CEM(+)	*3(+)	9.89282
196737	4020574	3	CEM(-)	*3(+)	8.07308
197186	2666984	3	CMUP(+)	3(+)	2.05675

Run	Event	Jets	Lepton (Charge)	Tag Level (Charge)	P_T (GeV/c)
197289	12158082	3	CEM(+)	3(-)	3.46003
198117	18873352	4	CEM(+)	*2(+)	16.0441
198623	4355115	4	CMUP(-)	3(+)	11.21
198695	2143563	3	CEM(+)	2(+)	2.31346
198695	10935618	4	CEM(+)	2(-)	14.5013
199187	4627063	3	CMXMK(+)	2(-)	2.82985
199218	512881	3	CMUP(-)	*1(-)	16.0742
199620	8673191	4	CMUP(-)	3(-), 2(+)	13.953, 8.14508
200051	3617608	3	CEM(+)	3(+)	2.72515
200756	6847547	3	CEM(+)	2(+)	9.80921
201133	7941729	5	CEM(+)	2(-)	2.37516
202135	4227665	3	CMX(-)	1(+)	3.56273
202739	329663	3	CEM(+)	1(-)	21.3924
203055	2308747	3	CMUP(+)	2(+)	6.40568
203055	10366024	4	CEM(+)	*3(+)	18.524
204469	621060	4	CMX(-)	1(-)	3.76186
204643	57292	4	CEM(+)	3(+)	2.95468
204650	2484068	4	CMUP(-)	*2(-)	2.46529
204679	2519459	4	CEM(-)	1(+)	7.0051
204750	4333547	4	CMUP(+)	3(-)	19.2292
205018	3180321	4	CMX(+)	2(-)	2.04222
205075	2784873	3	CEM(+)	3(-)	6.17981
205991	1000375	5	CEM(+)	1(+)	3.10252
206174	7593770	4	CEM(-)	1(+)	10.8167
206174	8328144	3	CMX(+)	3(-)	6.20788
206326	4815862	4	CMUP(-)	3(+)	32.0106
206326	6735620	4	CEM(-)	1(-)	2.57214
206537	15465618	3	CMX(-)	1(+)	8.96038
207079	2920112	4	CMUP(+)	2(+)	2.28691
207488	3747131	4	CEM(-)	2(-)	11.1663
209265	6003601	5	CEM(+)	2(-)	7.24506
209819	9422832	3	CEM(-)	2(-)	4.18177
211198	11680419	3	CMUP(-)	2(-)	3.19559
211292	1019493	5	CMX(+)	3(+)	2.40058
211441	1779903	3	CMUP(-)	3(+)	4.6627
219385	18776150	5	CEM(+)	3(-)	3.47815
219471	13586762	4	CEM(+)	2(+)	4.26714
219526	1311560	3	CMXMK(-)	2(+)	4.23584
220183	117469	4	CEM(-)	1(-)	7.60698
220221	9170416	3	CEM(-)	1(+)	5.65394
220247	3756690	5	CMUP(+)	3(-)	2.54441
220730	202366	4	CEM(-)	*3(+)	9.42665
221168	1324899	3	CEM(-)	*3(+), 3(-)	7.59146, 5.3019
221827	20625557	5	CEM(+)	1(-)	24.5628

Run	Event	Jets	Lepton (Charge)	Tag Level (Charge)	P_T (GeV/c)
222530	4706480	3	CMUP(+)	*2(+)	13.77
222552	12638633	4	CEM(-)	3(-)	59.6392
222576	6708173	5	CMUP(+)	3(-)	9.05422
222866	13564700	4	CEM(+)	2(+)	3.02753
222888	444625	3	CMUP(+)	2(-)	8.87015
223158	10007913	5	CEM(-)	1(-)	6.777
223309	7947707	3	CMX(-)	3(+)	7.4129
224161	9719229	3	CEM(+)	3(+)	4.45607
226246	1137424	3	CEM(+)	3(+)	2.38442
227554	5475443	4	CEM(+)	3(-)	4.64028
228536	11345511	5	CEM(+)	*2(+)	2.53436
228664	28203479	3	CMX(-)	1(+)	2.22495
228734	285480	3	CMUP(-)	3(-)	16.983
228738	4958511	3	CEM(+)	3(+)	16.9783
229534	3588204	4	CMUP(-)	3(-)	3.76741
230058	2236794	4	CMUP(+)	*3(-)	2.54778
230568	1147142	3	CEM(-)	3(+)	2.5476
230628	19303698	3	CEM(-)	1(+)	2.69236
230782	2792362	4	CMUP(+)	3(-)	2.4335
231060	1680648	4	CEM(-)	*3(+)	6.43139
231082	5537431	3	CEM(-)	*3(-)	6.65584
231084	7743139	4	CMUP(+)	3(+)	14.5337
231203	17948483	3	CEM(+)	1(+)	4.34726
231289	2388542	5	CEM(+)	3(-)	8.05151
231298	564062	4	CMUP(+)	3(-)	2.28765
231781	2126174	4	CMX(-)	1(-)	2.28387
231927	8754610	5	CMX(+)	3(+)	3.69979
231973	8536711	5	CEM(-)	3(+)	7.01965
232062	20171493	3	CEM(-)	1(-)	3.62514
232062	33333573	3	CMX(+)	1(-)	40.3665
232088	13469724	3	CEM(-)	2(-)	4.74655
232286	12638321	4	CEM(-)	3(+)	2.19568
232286	19645555	5	CEM(-)	3(+)	8.03703
232444	10228485	3	CEM(-)	1(+)	3.82534
233134	13047752	3	CEM(+)	3(-)	6.12649
233656	9770686	4	CEM(-)	2(+)	4.32426
234054	35475	3	CMX(+)	3(+)	6.15068
234572	6852934	3	CMUP(+)	2(-)	2.86846
234985	6221337	4	CEM(-)	*3(+)	7.48403
235983	236415	4	CMUP(+)	3(-)	2.82262
236255	22788664	3	CMUP(-)	*3(-)	2.8704
236762	1484347	3	CEM(+)	3(+), 3(+)	12.0611, 2.37338
236780	15141593	4	CMUP(+)	3(-)	7.25841
237522	1186946	3	CEM(-)	2(+)	17.7406

Run	Event	Jets	Lepton (Charge)	Tag Level (Charge)	P_T (GeV/c)
237705	25774420	3	CEM(+)	*2(+)	11.5965
237705	1663660	3	CMUP(-)	2(+)	4.09993

Chapter 10

Conclusions

We have made the first measurement of the $t\bar{t}$ cross section using soft electron tags in Run 2. Using 1.7 fb^{-1} of data, we have measured a cross section of

$$\sigma_{t\bar{t}}(\text{pb}) = 7.8 \pm 2.4(\text{stat}) \pm 1.6(\text{syst}) \pm 0.5(\text{lumi}), \quad (10.1)$$

a value consistent with the Standard Model expectation, $\sigma_{t\bar{t}} = 6.7 \pm 0.8$ assuming $m_{top} = 175 \text{ GeV}/c^2$. This measurement validates the SLT_e tagger, and demonstrates its usefulness for high p_T analyses at CDF.

The measurement of the cross section with this method is currently statistics limited, but only with twice as much data the statistical and systematic uncertainties will be comparable. As of September of 2008, CDF has already exceeded that total, recording approximately 4.0 fb^{-1} to tape.

The dominant systematic uncertainties are associated with the tagger itself. In particular, the data-MC conversion over-efficiency scale factor described in Chapter 5 accounts for a considerable fraction of the total systematic uncertainty. Removing this uncertainty would result in a 15% relative decrease in the overall systematic

uncertainty. When the tagger was being developed, the dominance of this uncertainty was unanticipated, and little effort was placed into reducing it. Understanding the p_T dependence of the over-efficiency scale factor which drives the large uncertainty would improve this measurement.

However, the primary reason for the large tagger uncertainty is the so-called Method 2 assumption, that is, whatever remains unaccounted for at the pretag level in the lepton+jets sample - after having accounted for MC derived backgrounds, QCD, and $t\bar{t}$ signal - is W+Jets. This assumption implies that if the effective tagging efficiency of W+Jets and $t\bar{t}$ are the same, then the method cannot distinguish between the two, and the uncertainty becomes unconstrained. Although the SLT_e tagger is more sensitive to $t\bar{t}$ than W+Jets, the large tagging uncertainty could be reduced significantly if a different W+Jets background estimation technique is employed.

As a comparison, the top cross section with the SLT_μ tagger is

$$8.7 \pm 1.1(\text{stat})_{-0.8}^{+0.9}(\text{syst}) \pm 0.5(\text{lumi}), \quad (10.2)$$

using a dataset 18% larger. Not only is the per track SLT_μ efficiency nearly two times higher, but the S/B is also lower. However, the SLT_μ does not perform better than the SLT_e for all analyses. At high p_T , the SLT_e has a higher purity than the SLT_μ because of the suppression of conversions and the declining fake SLT_e tag rate in that region. For analyses where purity is a major concern, such as the measurement of the top charge, the SLT_e will actually perform better. A preliminary measurement of the top charge has pique interest into the usefulness of the SLT_e .

Another consideration for the future is the release of the $|d'_0| < 0.3$ cm requirement on taggable SLT_e tracks. This was done to reduce background from conversion

electrons and mis-measured tracks. However, with the hindsight of recent results at CDF, it would be prudent to develop the tagger without such a requirement.

Finally, it will be interesting to watch the development of electron identification tools at the LHC experiments. In events with an order of magnitude greater energy and density, electron id will increasingly be thwarted by overlapping hadronic backgrounds. Perhaps lessons learned here will be useful at the next energy frontier.

Bibliography

- [1] C. Hall. *Measurement of the Isolated Direct Photon Cross Section with Conversions in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV*. PhD thesis, Harvard University, 2002.
- [2] W. M. Yao et al. Review of Particle Physics. *Journal of Physics G*, **33**:1+, 2006.
- [3] Q. R. Ahmad et al. Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, **87**(7):071301, 2001.
- [4] M. Kobayashi and T. Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, **49**:652–657, 1973.
- [5] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev. D*, **2**(7):1285–1292, Oct 1970.
- [6] J. J. Aubert et al. Experimental Observation of a Heavy Particle *J. Phys. Rev. Lett.*, **33**(23):1404–1406, Dec 1974.
- [7] J. E. Augustin et al. Discovery of a Narrow Resonance in e^+e^- Annihilation. *Phys. Rev. Lett.*, **33**(23):1406–1408, Dec 1974.
- [8] M. L. Perl et al. Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation. *Phys. Rev. Lett.*, **35**(22):1489–1492, Dec 1975.
- [9] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, **39**(5):252–255, Aug 1977.
- [10] F. Abe et al. Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, **74**(14):2626–2631, Apr 1995.
- [11] S. Abachi et al. Observation of the Top Quark. *Phys. Rev. Lett.*, **74**(14):2632–2637, Apr 1995.
- [12] K. Kodama et al. Observation of tau neutrino interactions. *Phys. Lett. B*, **504**:218–224, Apr 2001.

- [13] The ALEPH Collaboration, the DELPHI Collaboration, the L3 Collaboration, the OPAL Collaboration, the SLD Collaboration, the LEP Electroweak Working Group and the SLD Electroweak and Heavy Flavour Groups. Precision electroweak measurements on the Z resonance. *Phys. Rep.*, **427**(5-6):257–454, May 2006.
- [14] S. N. Ahmed et al. Constraints on Nucleon Decay via Invisible Modes from the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, **92**(10):102004, Mar 2004.
- [15] G. W. Bennett et al. Measurement of the Negative Muon Anomalous Magnetic Moment to 0.7 ppm. *Phys. Rev. Lett.*, **92**(16):161802, 2004.
- [16] P. Gambino and M. Misiak. Quark mass effects in $\bar{B} \rightarrow X_s \gamma$. *Nucl. Phys. B*, **611**:338–366, 2001.
- [17] T. Aaltonen et al. Search for $B_s \rightarrow \mu^+ \mu^-$ and $B_d \rightarrow \mu^+ \mu^-$ Decays with 2 fb^{-1} of $p\bar{p}$ Collisions. *Phys. Rev. Lett.*, **100**(10):101802, 2008.
- [18] S. Heinemeyer et al. Precise Prediction for M_W in the MSSM. *J. High Energy Phys.*, **06**(08):052, 2006.
- [19] The ALEPH Collaboration, the DELPHI Collaboration, the L3 Collaboration, the OPAL Collaboration, the LEP Working Group for Higgs Boson Searches. Search for the Standard Model Higgs boson at LEP. *Phys. Lett. B*, **565**(1):61–75, 2003.
- [20] M. Cacciari et al. The $t\bar{t}$ cross-section at 1.8 and 1.96 TeV: a study of the systematics due to parton densities and scale dependence. *J. High Energy Phys.*, **04**(04):068, 2004.
- [21] N. Kidonakis and R. Vogt. Next-to-next-to-leading order soft-gluon corrections in top quark hadroproduction. *Phys. Rev. D*, **68**(114014):1–15, 2003.
- [22] A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf. New extended model of hadrons. *Phys. Rev. D*, **9**(12):3471 – 3495, 1974.
- [23] J. Pumplin et al. New Generation of Parton Distributions with Uncertainties from Global QCD Analysis. *J. High Energy Phys.*, **02**(07):012, 2002.
- [24] B. W. Harris et al. Full differential single-top cross section in next-to-leading order QCD. *Phys. Rev. D*, **66**(054024):1–19, 2004.
- [25] V. M. Abazov et al. Evidence for Production of Single Top Quarks and First Direct Measurement of $|V_{tb}|$. *Phys. Rev. Lett.*, **98**(181802):1–8, 2007.

- [26] G. V. Jikia and S. R. Slabospitsky. Single top production at hadron UNK collider. *Phys. Lett. B*, **295**(1-2):136–142, 1992.
- [27] T. Aaltonen et al. Search for the Flavor Changing Neutral Current Decay $t \rightarrow Zq$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$. hep-ex:0805.2109, 2008.
- [28] D. Acosta et al. Study of the heavy flavor content of jets produced in association with W bosons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. D*, **65**(052007), 2007.
- [29] F. Abe et al. The CDFII Detector Technical Design Report. *Fermilab-Pub-96/390-E*, 2000.
- [30] F. Abe et al. The CDF Detector: An Overview. *Nucl. Instrum. Methods Phys. Res. A*, **271**(3):387–403, 1988.
- [31] C. S. Hill et al. L00: Operational experience and performance of the CDF II silicon detector. *Nucl. Instrum. Methods Phys. Res. A*, **530**(1-2):1–6, 2004.
- [32] A. Sill et al. SVX-II: CDF Run II silicon tracking projects. *Nucl. Instrum. Methods Phys. Res. A*, **447**(1-2):1–8, 2000.
- [33] A. Affolder et al. ISL: Intermediate silicon layers detector for the CDF experiment. *Nucl. Instrum. Methods Phys. Res. A*, **453**(1-2):84–88, 2000.
- [34] T. Affolder et al. COT central outer tracker. *Nucl. Instrum. Methods Phys. Res. A*, **526**(3):249–299, 2004.
- [35] L. Balka et al. The CDF central electromagnetic calorimeter. *Nucl. Instrum. Methods Phys. Res. A*, **267**(2-3):272–279, 1988.
- [36] S. Bertolucci et al. The CDF central and endwall hadronic calorimeters. *Nucl. Instrum. Methods Phys. Res. A*, **267**(2-3):301–314, 1988.
- [37] G. Ascoli et al. CDF central muon detector. *Nucl. Instrum. Methods Phys. Res. A*, **268**(1):33–40, 1988.
- [38] D. Acosta et al. The Performance of the CDF Run II Luminosity Monitor. *Nucl. Instrum. Methods Phys. Res. A*, **494**(1-3):57–62, 2002.
- [39] J. M. Campbell and R. K. Ellis. Update on vector boson pair production at hadron colliders. *Phys. Rev. D*, **60**(11):113006, 1999.
- [40] D. Acosta et al. First Measurements of Inclusive W and Z Cross Sections from Run II of the Fermilab Tevatron Collider. *Phys. Rev. Lett.*, **94**(9):091803, 2005.

-
- [41] D. Sherman. *Measurement of the Top Quark Pair Production Cross Section with 1.12 fb^{-1} of $p\bar{p}$ Collisions at $\sqrt{s} = 1.96 \text{ TeV}$* . PhD thesis, Harvard University, 2007.